An Investigation of the basis of representation of aesthetic objects held in collections in museums

A thesis submitted in fulfillment of the requirements for the

Degree of Doctor of Philosophy

Marian Reeves

B.Sc (Hons), M.Sc.

School of Computer Science & Statistics

University of Dublin

Trinity College

March 2019

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work.

I agree to deposit this thesis in the University's open access institutional repository or allow the Library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

Signed:__________________________

Marian Reeves, B.Sc. (Hons), M.Sc.

Summary

The design of robust databases for storing data about aesthetic objects is a major challenge, as manifested through discussions on metadata standards about these objects. These databases play a key role in the visualisation and conservation of cultural heritage, for instance. The access to such databases from a variety of users presents a significant opportunity for a number of stakeholders, including the general public, policy makers and art gallery curators.

A review of the literature highlighted the complexity of aesthetic objects and revealed that in recent years, there has been an increasing interest in artist archives as potentially fertile resources and raw materials for the art historian and curator. Furthermore, a recent innovation shows that integrating archives with artworks has the potential of uncovering new knowledge. This strong connection between artist archives and aesthetic objects motivated this study and the design of a prototype to enhance support for art research.

A mixed-methods study comprising a case study and a survey afforded the opportunity to gain insight into the curator conceptualisation of art and the information needs of art specialists. The case study included interviews with 4 curators (45 years experience), meetings and observations at the gallery, together with an examination of the body of knowledge continually evolving since the 1970's, including metadata and content standards and thesauri, providing essential knowledge for this study. A survey of 48 art specialists (with a combined total of 806 years experience) representing 20 museums in 9 countries across Europe, the USA and Canada, corroborated findings from the case study. This study revealed that domain experts responsible for the preservation of collections in their care, consult diverse resources when conducting their research and they (94% of respondents) believe that artist archives comprising unique, private and contemporaneous materials provide valuable background and contextual information for their research.

Established methods of conceptual modeling and logical modeling were used to create database schemas founded upon standards in the domain and a program was written to harvest archives from source texts to automatically populate the database. Population of the databases with paintings and their associated metadata and relations employed SQL scripts and scanned images for both the paintings and archival materials (comprising letters, sketches, artist clothing, photographs and miscellaneous ephemera) were uploaded to the system.

The prototype system combines aesthetic objects and artist's archives to provide a more comprehensive representation and visualisation than is typically available. A front-end GUI provides search and retrieval functionality and it allows the user to establish (reciprocal) relations between artworks and archives while at the same time capturing their reflections on these relations. The system was specified with the help of domain experts and tested with their help. Issues related to knowledge representation of aesthetic objects like ontological antecedents on one hand and provenance of such objects are also discussed.

Acknowledgements

I would like to thank my supervisor Professor Khurshid Ahmad, for his wisdom, guidance and continuous support throughout this challenging and fulfilling journey.

I would like to acknowledge the contribution of the many art experts to this study. From the National Gallery of Ireland, I would like to thank Dr. Adriaan Waiboer, Head of Research and Collections, for his advice and insight into the world of the curator and the arts domain; to Dr. Katherine Milligan, research curator; Leah Benson, archivist; Donal Maguire, Manager of Centre for the Study of Irish Art and to Raffaella Lanino and Catherine Sheridan for sharing their knowledge and their valuable insights into the domain. I would also like to thank Dr. Jill Unkel, Curator of Western Collection, at the Chester Beatty Library for her valuable complementary insights into the domain. Thank you also to all of the of the domain experts, worldwide, who gave their time to complete a survey, which informed the design of the prototype system for this research.

A special thank you also goes to Professor Brendan Tangney for his valued support and for including me in the CRITE activity lists and a long overdue thank you also to Dr’s Claire Conneely and John Lawlor for access to the Bridge 21 facilities.

A huge thank you goes to Shane Finan and Tatiana Macklin for sharing their knowledge of art with me and for their advice, time and feedback along with many stimulating conversations on campus, and during the all too infrequent visits to the golf course with Tatiana. Special thanks also to Dr. Doireann Wallace for her time to review and feedback on the survey, along with several computer science colleagues.

I am grateful to all my friends and colleagues, especially Aimee Borda, Artur Oliveira Gomes and Zeyan Zhao who transformed our office into a very welcoming and positive environment and to Dr’s Jason Cook, Stephen Kelly, Xiubo Zhang and Aibhin Bray for their support, friendship and many social sessions and coffee breaks.

I would also like to thank the School of Computer Science, at Trinity College Dublin for generously funding my attendance at a Protégé workshop at Stanford University, California.

Finally, and most importantly, thank you to my husband Denis, for his love, support and encouragement throughout my PhD studies and who made all of this possible.

I dedicate this dissertation to my parents, Mary and Tony O’Rourke, who would have been so proud.

Contents

List of Figures

List of Tables

1. Introduction

According to the most recent estimates there are more than 55,000 museums in 202 countries of the world[1]. Indicating the size and coverage of people in the cultural heritage domain, two of the several umbrella organisations within the domain published the following information: The International Council of Museums (ICOM) lists 35,000 members and 20,000 museums. The Association of Art Museum Curators (AAMC) in America lists 1,300 members from over 400 institutions located in 18 countries.[2] As not-for-profit permanent organisations, ICOM defines the role of museums as serving society through preservation, conservation, education and promotion of collections of tangible and intangible cultural and natural heritage. Museum collections comprise a whole spectrum of aesthetic objects from fine art paintings, to manuscripts, to sculpture, to prints, to drawings and photography. Literally millions of these artefacts and warehouses of archives containing collectible objects from sketches, to paintbrushes, to clothing, to diaries, to letters and miscellaneous ephemera reside in museum holdings throughout the world.

1.1. Motivation

Within museums, domain experts, curators and archivists / curators of archives, are responsible for the preservation of collections in their care. As content specialists, curators possess knowledge of their collections and develop new insights through continued research and as new collateral materials become available. Perhaps an obvious manifestation of their specialist knowledge lies in exhibition curation. Drawing upon their specialist knowledge, curators purposely select works of art and purposely arrange the works of art and collateral texts in such a way that reflects their interpretation of the works and the exhibition theme. Through these decisions, combined with attention to lighting and space, curators exercise considerable influence on the visitor's experience at the museum (Moser, 2010; Tröndle, Greenwood, Bitterli, & van den Berg, 2014; Tzortzi, 2014).

This study focuses on visual cultural heritage, specifically, on aesthetic objects such as paintings, held in collections in museums. The principal question examined is how a computational representation can incorporate the knowledge of curators and support scholarly research in art.

Creating a 'good' representation (Winston, 1992, p. 18) of curatorial knowledge presents an interesting challenge, especially given the views expressed in the philosophical literature (Lamarque, 2012; Margolis, 1977; Thomasson, 2004; Walton, 1970) highlighting the complex and heterogeneous nature of aesthetic objects. The philosophical literature also emphasises

1 http://icom.museum/resources/frequently-asked-questions/ Last viewed 08/01/2018
2 http://www.artcurators.org/?page=AboutUs Last viewed 08/01/2018

the need for ‘grounders’ or domain experts, and their art specialist knowledge and practices (Gombrich, 1995; Panofsky, 1972; Wölfflin, 1932) to uncover the multifaceted knowledge to interpret and evaluate works of arts. Although the work of art and its intrinsic qualities remain central to their research, art specialists consider external information essential towards enhancing their understanding of the authentic nature of the work. Hence, there is increased interest in archival materials, particularly artist archives.

Artist archives, comprising unique, private, contemporaneous materials generated over the course of the artist’s lifetime, represent potentially fertile resources or raw materials for the art historian and curator (Breakell, 2015; Pyle, 2001; M. Reed, 2017; Sjöholm, 2014). According to M. Reed (2017) artist archives, combined with other public art historical resources “are foundational to creating the canon of the most valued or revered works of art and artists”. Moreover, the authors also consider archives or impending archives of contemporary artists’ studio materials, important sources of inspiration for the artist himself/herself and they note their importance as sources of documentary evidence of installation art long after it has been dismantled. Koot and Koppenol (2016) note an increased interest in ephemeral materials in art historical research, representing rich resources for information, especially about relationships between artists, dealers and museums.

Figure 1.1 The “The Liffey Swim” (centre), a painting by Irish artist Jack B Yeats (1871-1957) winner of the Silver Olympic Medal (left) in the Arts & Culture section, at the 1924 Paris Olympics. Also shown, is a Register of Awards (top right) and a page from a Register of Paintings (bottom right) from the Yeats archives, at the National Gallery of Ireland

Consider, for instance, the artist Jack B Yeats (1871-1957), arguably Ireland’s greatest artist of the 20th century. “The Liffey Swim” (Figure 1.1, centre) painted by Yeats in 1923, celebrates the Liffey Swim, an annual event held in Dublin city, since the inaugural swim on 22 July 1920 and was awarded the Silver Medal at the 1924 Paris Olympics, in the 'Concourse d'Art section (Culleton, 2014; K. McCarthy, 2012). To examine the image one needs to have access to the archives. The artist’s archives, housed in the National Gallery of Ireland comprise hundreds of artefacts; from sketchbooks, to letters, to the artist’s paint brush, palette, smock and easel, to photographs, to miscellaneous ephemera and includes the Olympic Medal (Figure 1.1, left), the

Register of Awards (Figure 1.1, top right) and a page from a Register of Paintings (Figure 1.1,bottom right) an alphabetically arranged, handwritten record, of the artists paintings. Each of these elements provides contextual information that helps the curator to understand the artist, influences, inspiration and intention of the artworks, and these need to be captured in the design of a database (more on this in 0 and 0).

1.2. Context

Given the technologies available today, it is difficult to imagine that until the 1960's museums recorded information in paper-based records (Marty, Rayward, & Twidale, 2003) and catalogue cards containing skeletal information, used to search and discover what artefacts existed in museum holdings. Search and retrieval remain as relevant today as it did at that time. Of greater consequence, however is the fact that at that time, the knowledge in museums resided in the works of art and in the minds of curators and their paper-based files (Ibid) with limited access to either one.

The emergence of relational database management systems in the 1970's (Codd, 1970; Elmasri & Navathe, 2014), available commercially in the 1980's, transformed museum operations. An immense body of knowledge contributing to computational representation, administration and classification of the diverse repositories in archives and museums (Figure 1.2 , Bottom section) has accumulated since the 1970's. Vocabulary development, for semantic annotation of cultural heritage materials is ongoing since the late 1970's by the Getty Research Institute (Figure 1.2 Top section), to meet the needs of the arts community (Baca, 2004; Harpring, 2010a; van Hooland et al., 2013; Weinberg, 1995). These vocabularies promote consistent cataloging and metadata content, supporting search and retrieval. Similarly, work continues on classification systems (Figure 1.2 second section) for iconographical analysis and description of images or works of visual art, including Iconclass (Couprie, 1978; VanStraten, 1986) and the Index of Christian Art at Princeton University (Hourihane, 2014). In the 1990's, work commenced on metadata standards (Figure 1.2, last section) specifically for the cultural heritage domain (Baca, 2008a; Baca & Harpring, 2014 -Revised online 2016; Baca, Harpring, Lanzi, McRae, & Whiteside, 2006; Doerr, 2003; Pitti, 1999; VRA, 2007; Weibel & Lagoze, 1997). Adopting appropriate metadata standards not only contributes to representation consistency, but also to resource discovery and retrieval; shareability and interoperability.

Vocabularies	
ULAN	Union List of Artist Names
AAT	Art and Architecture Thesaurus
TGN	Thesaurus of Geographic Names
Classification	
Iconclass	Classification systems for Iconographic Analysis
Index of Christian Art	Index of Christian Art at Princeton University, Classification system for Iconographic Analysis
Metadata Schemata	Name and Description
DC	Dublin Core: Used for simple and generic description of electronic resources.
VRA	Visual Resources Association: Describes digital Images and works of Art and Culture
EAD	Encoded Archival Description: Atandard for archival and manuscript repositories.
SPECTRUM	Standard ProcEdures for CollecTions Recording Used in Museums: Catalogue & manage materials in libraries, archives and museums (LAMs).
CDWA &	Categories for the Description of Works of Art: Describes works of art, architecture, other material culture, groups and collections of works, and related images
CDWA Lite	Reduced set of CDWA categories
ISADG	General International Standards for Archival Description.
CIDOC CRM	Conceptual Reference Model: High Level Ontology for CH materials in LAMs
EDM	Europeana Data Model: Designed for the Europeana project, describing semantics of materials in LAMS, in Europe.

Figure 1.2: of knowledge for cataloging, annotation and classification in museums and archives.

With the introduction of the world wide web in the 1990's museums began to publish their collections online by the end of the century and content on the web pages is continually updated, as observed from viewing museum web pages. Motivation for research into computational ontologies gained momentum at that time also, as part of an effort to reduce duplication of effort and to maximise knowledge sharing and systems interoperability (Gruber, 1993, 1995; Neches et al., 1991). With the emergence of the semantic web (Berners-Lee, Hendler, & Lassila, 2001) technologies offered new formats and structures for domain representation and computational ontologies began to play a prominent role in some domains, in the twenty first century. Ontology based systems enable common interpretation and consistent usage of domain knowledge across user communities achieved through an agreed vocabulary of terms organised in a way that describes the semantics of the domain, and to facilitate automatic inferences across the ontology. Computational ontologies have been adopted in a range of medical domains (Ivanović & Budimac, 2014) and biomedical domains, to standardise terminology and to provide access to domain knowledge (Hoehndorf, Dumontier, & Gkoutos, 2013). Representing a domain using ontology based systems proves to be a complex task, however and evaluation of some systems reveals the need for continual refinement and testing to eliminate anomalies and ensure accurate and comprehensive representation of the domain (Campbell, Campbell, West, McClay, & Hinrichs, 2014; Rector, Brandt, & Schneider, 2011).

Evidence of research into ontology based systems undertaken in the cultural heritage domain, includes CLAROS and Europeana. The Classical Art Research Online Services (CLAROS), an international, interdisciplinary research project led by Oxford University (Kurtz et al., 2009) implemented their system using an extended version of the CIDOC-CRM ontology, launched in 2011. The Europeana project describes their system as an ontology, namely, the Europeana

Data Model (EDM). Europeana ingests and aggregates artefact details and images from participating cultural heritage institutions and provides a web portal for the discovery of these objects and according to Heffernan (2015), Europeana may enhance cross-cultural studies. Research based on collections at the Smithsonian American Art Museum also adopted the EDM conceptual reference model (Szekely et al., 2013).

By 2016, however, the landscape of the semantic web was changing. Considering the vast volumes of distributed, heterogeneous datasets on the web, from formal ontology based systems; to linked open data (LOD) that connects to other linked data through URI's (Uniform Resource Identifiers) to web pages using standard data mark-up schema; to streaming datasets, Bernstein, Hendler, and Noy (2016) contend that the challenge for the next decade will be how best to leverage these datasets for intelligent applications. Moreover, the authors observe that perhaps more appropriate methods that reflect how data is actually used, lie somewhere between the formal ontology approach at one extreme of the spectrum, to unstructured and less controlled data at the other. They propose that research into "more meaningful services that rely less on logic-based approaches and more on evidence-based approaches", is required.

In contrast to centralised repositories, ontologies and to LOD, to access Fine Art information, recent research investigates using API's (Application Programming Interface), that provide search and retrieval of the museum's online catalogues (Buchanan, Capanni, & González-Vélez, 2015). The authors propose that this alternative approach provides access to up-to-date information from trusted sources, through intelligent applications that automatically search participating institution's repositories, precluding the need to search individual websites and for centralised repositories. The API approach suggests a pragmatic and promising solution for curators, who require the latest, reliable information, compiled by domain experts, when they are researching a work of art.

A recent innovation in the cultural heritage domain, RKD Explore, (Koot, 2015, p. 1000) presents a web portal for art historical research that incorporates 7 databases, comprising Artist, Image, Library, Collections, Portraits, Excerpts and the recently added Archives database. The author reports that database development at RKD (The Netherlands Institute for Art History) began in 1990 with the RKD Artist database only, and observes that the technology the databases were built on remains largely the same. Strengths of the relational database model, for representing the complexities of the cultural heritage domain, include design principles enabling multiple and diverse relations between the different entities in the domain and data abstraction through multiple views to aggregate, combine or define a subset of the data from the underlying base tables.

Meanwhile, global deployment of relational database systems (RDB) in industry, in business critical applications and in the cultural heritage domain [3] continues (Table 1.1). These systems, deployed in thousands of institutions worldwide, are validated as SPECTRUM compliant by the Collections Trust UK. The databases are either web based, as the backbone of museum web applications and/or represent the source repository for data extracts uploaded to online systems and union catalogues. In addition to enabling diverse relations among entities, further advantages of the RDB include secure management and efficient organisation, storage, and search and retrieval, of large volumes of diverse data, as can be found in the cultural heritage domain.

Table 1.1 Cataloging systems for the Cultural Heritage domain: details from the Collections Trust UK

	Software	Museum	Archives	Library	RDB	Storage Model
1	Adlib	✓	✓	✓	✓	MS SQL Server
2	Proficio	✓	✓	✓	✓	MS SQL Server
3	Emu	✓	✓	✓		ODB
4	Mobydoc	✓	✓		✓	MS SQL Server
5	MuseumPlus RIA	✓	✓		✓	MS SQL-Server, PostgrSQL, MySQL, Oracle
6	Primus	✓	✓		✓	Oracle 11G or newer.
7	Qi	✓	✓		✓	MS SQL Server
8	Sofie	✓	✓		✓	MySQL.
9	The Museum System (TMS)	✓	✓		✓	SQL Server 2012/14 Oracle 11gR2 or 12c
10	Vernon CMS	✓	✓		✓	RDB-OpenInsight (Linear Hash DBMS)
11	Modes Complete	✓	✓			NexusDB database
12	MuseumIndex+	✓	✓			NoSQL
13	Robotron*Daphne	✓			✓	Oracle RDBS
14	S-Museum	✓				SQL solution storing as key/value & ODB
15	eHive		✓	✓	✓	MySQL.
16	Calm		✓			DScribe, proprietary text retrieval &relational sys

1.3. Aim of this Research

The aim of this research is to incorporate the knowledge of museum curators in a computational representation that may support scholarly research in the domain. To achieve that aim this study investigates how schemata designed for the systematic organization of aesthetic objects and the relationships of these objects to other primary sources, of archival materials, may be combined to facilitate a more comprehensive representation of curatorial knowledge than is typically available. To advance these aims the following question and its sub questions are posed:

How can a computational representation of a complex and knowledge rich domain support research in that domain leading to enhanced intellectual value?

Sub questions:

1. *What kinds of information are considered by art specialists when researching a work of art?*

[3] http://collectionstrust.org.uk/software/ last visited 25/02/2018

Art specialists possess considerable training and expertise not only of the domain, but also how to search the domain itself. They can rely on small items of data to launch a comprehensive search and are able to evaluate retrieved data for its precision and recall. The quality of the retrieved data, however, depends on the consistent application of appropriate metadata to store the data and relations of interest to curators, and to support systems interoperability in order to locate and retrieve related materials.

To answer this question the researcher adopted an exploratory mixed-methods research approach comprising a case study followed by a survey (1.4.1)

2. *How can database design principles support an enhanced representation of art that may underpin future intelligent applications?*

Given the rise of large data stores it is possible to store structured data and unstructured data in ever increasing volumes. Art and other heritage establishments typically store highly structured data about their holdings, comprising coded information about objects, location, collection, artists, medium of production and these systems are essentially inventory control systems. More recently, unstructured information like excerpts from a file or a file itself have been used to augment the information in the structured components and ad hoc links established between the two based on expert knowledge and ability to search the domain.

To answer this question, methods of database design incorporating standards developed for the cultural heritage domain, were employed (1.4.2).

3. *How can the system inform and support domain expert's research process?*

This perhaps is a complex question and an ambitious one. The research shows that a prototype application, comprising a front-end graphical user interface, underpinned by the representation, may support a heritage researcher by incorporating some heuristic of searching the domain as incorporated in this study. More information is provided on this (1.4.4)

1.4. Research Approach

The primary aim of this study is to design a computational representation that captures curatorial knowledge. Achieving that aim requires an understanding of the ontological commitment of the curator; that includes the role of the curator and the kinds of things and their relations, they consider important when researching a work of art. For that reason, and having considered other art institutions, we approached experts at the National Gallery of Ireland (NGI)

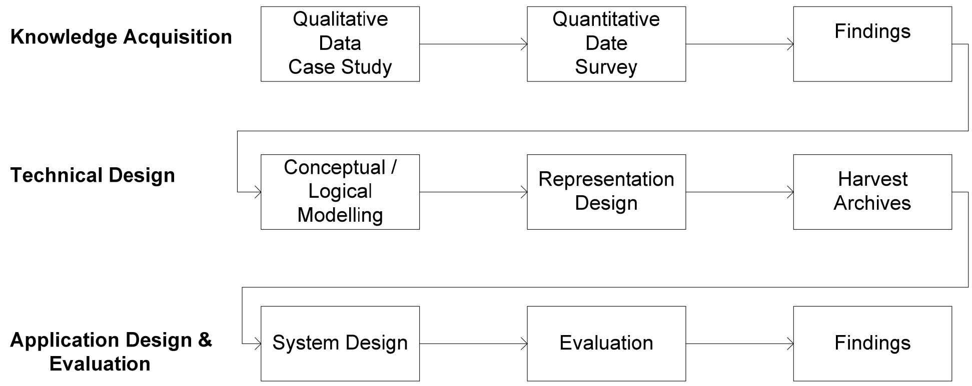


Figure 1.3: Research Design: A 'blueprint' for the logic that links methods and collection and analysis of relevant data to answer the research question (Yin, 2015, pp, 27-29)

Consequently, this research is carried out in cooperation with experts at NGI, in that it is informed by inputs from curators, an archivist, a collections management administrator and a research curator, over the course of many interactions with these experts. Also, including a curator from the Chester Beatty Library, a designated National Cultural Institution, offers a complementary perspective.

The research design (Figure 1.3) comprising an initial exploratory qualitative case study, fit very well with this study, providing an opportunity to gain insight into important aspects of the domain, especially regarding curatorial practices and information needs. Subsequently, a survey tested the prevalence of case study finding over a larger population and it engendered a deeper understanding, from a variety of perspectives (Q1). Findings from the mixed methods study resulted in refinement of the research question and informed the design of the computational representation. Designing the representation involved database design techniques and methods of text analysis and harvesting data from texts (Q2). Finally, evaluation of the extent to which the representation captures curatorial knowledge and may contribute to scholarly research comprised demonstration of the system to domain experts and their feedback on the system, when researching a work of art (Q3).

1.4.1. Mixed Methods Study – Knowledge acquisition

To answer Q1: *What kinds of information are considered by domain experts when researching a work of art?* , a mixed methods study was undertaken.

Mixed methods research is increasingly adopted by the information systems research community as a means to understand the user of technology and its value and impact on users (Venkatesh, Brown, & Bala, 2013). This research approach uses both qualitative and quantitative research techniques and methods that share the same questions, enabling collection of complementary data (Creswell, 2014, pp. 217-219; Yin, 2014, pp. 65-67). The authors contend that data collection from these differing perspectives strengthens the validity of the research and the inferences that may be drawn from it. This research commenced with a case study, which informed the design of a survey questionnaire.

A case study is an empirical enquiry that investigates a contemporary phenomenon within its natural setting, especially when the boundaries between phenomenon and context may not be clear (Yin, 2014, pp. 14-17). Moreover, unlike in laboratory or field experiments, case study researchers typically have less a priori knowledge of the specific variables of interest and they do not seek to control or manipulate variables under study (Ibid.) Undertaking an exploratory case study provided insight into the nature and complexity of the phenomenon, that being, the curators role and their conceptualization of works of art and how they develop knowledge of works of art (Ibid.;Creswell, 2014, p. 40). In seeking to understand these aspects of the domain, the case study explored:

- the role of the curator
- how primary sources of artist's archives contribute to art research
- Information and the sources considered relevant to scholarly research of art

Case study data sources comprised semi-structured interviews with 4 domain experts, observation, fieldwork and examination of documents and standards developed for the domain. The case study findings assisted in refining the research question (Rai, 2017; Yin, 2014, p. 238) and improved this researcher's understanding of the domain, however, conducting a survey provided alternative perspectives and tested the prevalence of views expressed, among a larger population.

Of the many advantages of web-based surveys (Bryman, 2012, pp. 673-678), specifically for this study, it provided a more economical access to a larger population and it facilitated rapid turnaround in data collection; data were stored directly into a secure database. On the other hand, careful questionnaire design and pre-testing to remove ambiguity and promote consistency of interpretation across the population, was imperative. When designing the survey the researcher considered the purpose of the survey, whether to use a probability sample or not, the sample frame, the sample size, questionnaire design and response rates (Creswell, 2014, pp. 155-159; Fowler Jr, 2014, pp. 6-11). The sample frame was restricted to curators and domain experts who carry out research on works of art. The sampling procedure was random in so far as the researcher randomly browsed museum websites, trawling for curator names and email addresses. Of the 200 emails sent to prospective expert participants, 60 individuals responded, yielding 48 valid responses from art experts based in different museums, in nine different countries.

1.4.2. Database design – representing curatorial knowledge

To answer Q2: *How can database design principles support an enhanced representation of art that may underpin future intelligent applications?* methods of database design were employed.

Modelling the real world has been addressed by database experts since the 1960's where they abstracted important data using methods of conceptual modelling (Bachman, 1969; Chen, 1976). The conceptual modelling method employed in this research comprised Entity Relationship (ER) modelling (Chen, 1976) and Extended Entity Relationship (EER) modelling (Teorey, Yang, & Fry, 1986). Used during top down analysis, and based on findings from the mixed methods study, the model illustrates aspects of reality in graphical form (Chapter 3), using constructs of *entity types*, *relationships* and *attributes* and extended to include *superclass/subclass*, *category* and *aggregation*.

Mapping from the conceptual model to a logical model, captured aspects of reality at a more detailed level in a database schema and employed established procedures and techniques (Elmasri & Navathe, 2014, pp. 287-301; Teorey et al., 1986). Incorporating fundamental principles of relational database design (Codd, 1970), of functional dependencies and normalization, contributed towards a well structured schema. (Elmasri & Navathe, 2014, pp. 520-543; Fagin, 1977, 1981; Kent, 1983), comprising a set of minimally defined base relations, while at the same time establishing different kinds of relationships among the entities. Through these relationships the database facilitates access to the diverse information to describe the complex objects in the domain. Furthermore, by incorporating metadata and content standards developed for the domain, the representation includes appropriate entities and attributes to describe these objects and their relations.

1.4.3. Data collection and analysis

The qualitative and quantitative data sets were analyzed separately (Creswell, 2014, pp. 222-223; Yin, 2014, p. 193). The report from the mixed methods study synthesizes the methods, techniques, and concepts and themes arising from the integration of both the qualitative and quantitative data sets (Johnson, Onwuegbuzie, & Turner, 2007). Qualitative data were stored in the research database, albeit a set of excel spreadsheets and word documents and exporting the survey data from the online database to Excel spreadsheets facilitated statistical analysis of the quantitative data and qualitative analysis of responses to open questions.

Directed and indirect content methods were used on the qualitative data (Bazeley, 2013, pp. 125-186; Hseih & Shannon, 2005) to test initial hypotheses and to uncover concepts and themes emerging from the data, and quantitative data analysis used statistical methods. Furthermore, triangulating the data from all sources in the both methods determined the level of consistency or corroboration of the findings (Sekaran & Bougie, 2016, p. 106; Yin, 2014, pp. 120-121)

1.4.4. Evaluation – application design and evaluation

To answer Q3: *How can the representation be used to inform and support domain expert's research process?*

Based on formative evaluation from domain experts, the researcher designed a prototype application incorporating search and retrieval functions, for both works of art and archives. The application automatically retrieves related archival materials when a work of art is retrieved and displayed. Curators can link archival materials to a work of art and record their interpretations and reflections at the same time. This knowledge is available for sharing among the art expert user community

Evaluation of the application involved researcher queries and the contribution towards capturing curatorial knowledge and supporting curatorial research was carried out through presentation and demonstration of the system on a one-to-one basis with 3 domain professionals.

1.5. Main findings

Findings from the mixed methods study (that included 4 interviews and 48 survey responses from domain specialists) show that domain experts responsible for the preservation of collections in their care, consult many and diverse materials to develop their knowledge of the collections. Provenance, publications about the artist and works of art and location of an artwork, are of primary interest to these experts. The study highlights, however, the strong connection between artist archives and works of art, (69% of survey responses) slightly more so than between works of art (60%), showing that archives represent valuable research resources for the art specialists carrying out their art research. Moreover, when the question focused specifically on archives, 94% of survey respondents believe that artist archives can deepen understanding of the artist and the paintings. This shows that the contextual and complementary information from archives can alter perspectives when viewed through the lens of these unique, contemporary materials generated during the artist's lifetime.

The resources required by domain specialists are not always readily available in their museums systems and locating the materials and capturing these relations is an arduous task. Collections Managements Systems incorporate some of domain metadata, vocabulary and classification standards that promote consistency of description and facilitate search and retrieval and interoperability. Museums may store different collections (for example, works of art and archives), in the same database, although organizing both datasets according to different underlying schemata. That being said, some institutions manage their different collections in separate systems and separate databases.

Currently, establishing relations between objects, such as a painting and a document or a file, is enabled through metadata elements. However, employing this method assumes that the cataloger knows the exact reference; the URI or file path, of the related object, to facilitate its retrieval. Furthermore, the referred object is unaware of the relation between it and the original object, lacking an automated reciprocal relation in the referred object (although some standards recommend specifying reciprocal relations where possible). Therefore, unless the materials are displayed with the work of art and/or unless reciprocal relations are established, the curator may not always be aware of these relations. An alternative mechanism used is by a textual description in both object records. Inherent limitations of both of these methods however are, first, the assumption that the cataloger is sensitive to the connection between the materials, but also that using either method precludes capture of reflections from one or more curators, conducting their research on the same objects.

Finding from this study shows that curatorial knowledge can be incorporated in a representation, in support of scholarly research of the complex objects in their care. This research describes a framework for building a representation and an application that allows domain experts to establish relations between works of art and archives, stored in different databases, without requiring knowledge of specific URI's or file paths.

An intelligent application that draws together collections of aesthetic objects and archival materials from different repositories presents a more comprehensive representation and display of an artwork, than is typically available. Enabling curators to explicitly establish relations between these objects and to store their reflections about the relationship, provides an alternative method of capturing and sharing curatorial knowledge among the art specialist community.

1.6. Contributions

This research incorporates knowledge contributed by domain experts responsible for the preservation of works of art and archives and the prototype developed in this study, assimilates that knowledge and addresses aspects of curator's needs, when researching a work of art.

Finding from the mixed methods study, that included interviews with 4 art experts (and observations, documentation review and fieldwork), and survey responses from 48 art experts, revealed the kinds of information that curators consider important when researching a work of art, and the diverse resources that they consult. Findings revealed the strong connection between works of art and archival materials, highlighting the value that curators attribute to artist archives towards enhanced intellectual value of the work of art.

The main contribution of this research is the design of a system that brings together collections of aesthetic objects and archival materials, from different repositories, that automatically displays related archival materials when the user views an object. Currently, metadata schemata for the domain facilitate relationships between objects in the same database and to reference external objects, but unless the materials are displayed with the work of art and/or unless reciprocal relations are established, the curator may not always be aware of these relations.

The system also enables curators to explicitly create relationships between a work of art and archival materials and to store these relations in a separate database. The system provides:

1. A mechanism to establish (multiple) relationships between works of art and archival materials, stored in different databases, and to share this knowledge among the art expert community.

2. A virtual space that allows examination of a work of art, while also allowing the user to view associated artist archives. This presents another potential method for curators to carry out their research by providing complementary perspectives on a work of art; through the lens of contemporary materials generated during the artist's lifetime.

   Features of the system include:

   - Search and retrieval of works of art and archives that are held in separate databases.
   - Relations between works of art and archives can be established without knowledge of the underlying metadata schemata or the exact file path or URI of the object.
   - Capture curator reflections and comments on the relationship between a work of art and an item in the archives, when they establish the relationship between the two objects.
   - All previously established relationships are visible, including those defined by other curators, when viewing an object.

In this work we have looked at how 4 domain experts conceptualize works of art and these collaterals in curatorial work, which involves classification, display, and making the collection available to other experts, primarily through their classification schema. These experts were interviewed over a period of 4 years and this has led to the design of a representation of 50 works of art and 350 collateral items from the archives.

The system was demonstrated to experts who reviewed it relatively favourably.

The researcher also developed a program to harvest archival materials from text documents and to map the extracted data to a database. This program may be adapted to harvest data from different texts/metadata standards, to automate database population.

1.7. The structure of the thesis

Following the introduction presented in this chapter the remainder of this study is described in the following chapters.

Chapter 2 presents a review of the literature of the theoretical and practical research relevant to this study. Principles of knowledge representation and ontology provide the theoretical foundation and technologies for representation, are discussed. The complex nature of artworks is discussed along with the continually evolving body of knowledge in the domain for the representation of art. Recent research and system implementations are also reviewed.

Chapter 3 describes methods of database design, comprising conceptual modeling and logical modeling and details the method of harvesting archives from text documents. It also discusses approaches for knowledge acquisition and identifies a mixed methods design comprising a case study and survey together with data analysis and ethical considerations.

Chapter 4 presents an exploratory case study that provides insight into the conceptualization of artworks together with standards developed for the domain. This qualitative data tests the initial hypothesis and informs the design of a survey that provides a quantitative perspective and tests the prevalence of opinions across a larger population.

Chapter 5 describes the design of a prototype application that incorporates archival objects and works of art providing a more comprehensive representation than typically available and subsequent evaluation by art specialists / curators.

Chapter 6 concludes by drawing together all aspects of the study and the findings from the mixed methods study and the prototype evaluation. The contributions of this study together with the limitations and considerations for future research are presented.

2. Literature Review

This chapter presents a review of the literature relevant to the representation of complex and heterogeneous objects in a knowledge rich domain, such as the cultural heritage domain. Literature spanning knowledge representation, philosophy, art, computational representation models, representation standards and museum systems inform this study.

Key concepts underpinning this study, of representation, philosophical ontology and models for representation, including ontology as adopted in computer science, OWL, Topic Maps, OAI-PMH and relational databases, are examined (Section 2.1). Aspects of the ontology of art, art specialist practices, art research studies and primary sources, of interest to domain specialist conducting research of artworks, are fundamental to understanding and representing this complex domain (Section 2.2). Resources and standards for organizing, describing, annotating and classifying of works of art and collateral materials identifies, play an important role in the representation of aesthetic objects and are described next (Section 2.3). Web implementations in the cultural heritage domain are reviewed (Section 2.4). The chapter concludes with a discussion of the materials presented, highlighting the area for further research in this study (2.5).

2.1. Key Concepts

The primary aim of this research is to incorporate the knowledge of museum curators, responsible for the preservation of collections in their care, in a computational representation. Principles of knowledge representation, traditionally the focus of artificial intelligence are presented. Next, exploration of Ontology as the science of being, originating in a branch of philosophy known as Metaphysics, followed by its adaptation by computer science. The components of an ontology in computer science are described and a discussion of the influence of philosophical ontology on ontology in computer science. Consideration of an alternate technology, the relational database model, concludes this section

2.1.1. Knowledge representation

The concept of representation is considered by various disciplines in the literature. A broad view from a cognitive science perspective posits that the fundamental components of a representation comprise ‘a Represented World’ and the ‘Representing World’, where mapping rules between elements of the former to the latter world determine correspondence between both worlds (Markman, 2013, pp. 5-8). Furthermore, a representation necessitates some mechanism for using or reasoning over the representing world (Ibid).

From a computer science perspective the analogous term, knowledge representation, refers to the study of how what we know can be represented and reasoned with in a computational model, (Brachman & Levesque, 2004, p. v; Grimm, Pascal, & Abecker, 2007; Sowa, 2000, p.

132) that captures knowledge of a domain or task, in the real world. Winston (1992, p. 18) characterizes good representations as those that:

- explicitly define the important objects and their relationships
- express how one object or relation influences another
- bring objects and relations together
- suppress irrelevant detail (hide rarely used information until required)

Moreover, Winston asserts that good representations are transparent, complete, and concise and are computable, in that they can be created with an existing procedure and it enables rapid information storage and retrieval.

Davis, Shrobe, and Szolovits (1993) propose a broad framework for a successful knowledge representation (KR) comprising five essential roles played in artificial intelligence, further explicated by (Sowa, 2000, pp. 134-140).

1. *KR is a surrogate*: Real world objects, relations and processes must be represented in a computational model as symbols which can be understood and reasoned over. Correspondence between the surrogate and the real world is determined by the semantics defined in the surrogate

2. *KR is a set of ontological commitments*: This comprises the kinds of things (such as objects, events and processes), their properties, relations and particulars, purposely selected as appropriate to represent the task or domain under consideration.

3. *KR is a fragmentary theory of intelligent reasoning*: A fundamental aspect of a representation is reasoning. Procedural code and declarative statements provide different mechanisms to process or reason over a representation. The procedural approach defines a sequence of operations performed (comparable to knowledge-how) whereas the declarative approach (comparable to knowledge-that) presents a set of propositions. According to (Davis et al., 1993) the essential questions in selecting the appropriate approach must be "What can we infer from what we know?" and "What ought we to infer from what we know?" The authors conclude that "logic offers both a degree of generality and the possibility of making information about recommended inferences explicit and available to be reasoned about in turn".

4. *KR is a medium for efficient computation*: Striking a balance between the correspondence of a computational model to the real world and the effort and time to reason over the model is fundamental to a successful representation. Advancements in hardware, programming and representation languages present new opportunities for efficient representation.

5. *KR is a medium of human expression*: A representation expresses experts' domain knowledge, communicated to and captured or encoded by knowledge engineers in a

language understood by the technology employed. Moreover, it should be possible for domain experts to inspect and verify accurate representation.

Arguably, and Entity Relation Diagram (Chen, 1976) and especially the Extended Entity Relationship Diagram (Teorey et al., 1986) could be classed as representations, given that they capture an accurate representation, illustrated on graphical form. However, Sowa (2000) contends that although the entity relationship diagram (ERD) represents some of the information needed for a conceptual schema, such as the entity types and their relations and some of the relations constraints, it lacks the necessary finer detail for data definition and integrity constrains and cannot express the procedures to manipulate the data. Furthermore, the extended ER model cannot represent object encapsulation or express message passing or the operations and object performs in response to a message (Ibid). These observations show that the ER and EER are limited in their representation and do not possess all of the characteristics of a good representation, as previously defined by Winston (1992).

Similarly, database practitioners contend that a relational database should be classed as a representation. According to Winston (1992, p. 626) a relational database is a data structure; containing lexical, structural, constructor, reader and writer components, and that as such it is not a representation. However, Winston goes on to say that “It becomes a representation once you say something about the semantics involved in a particular relational database”.

A number of technologies have been employed for knowledge representation such as, frames, semantic networks, rules and logic: a subset of which are fundamental to ontology in computer science. Before examining representation models of ontology (as adopted by computer science) and relational databases, the term ontology is first explored from its original philosophical perspective.

2.1.2. Ontology and Philosophy

“What is there?” asks Quine, thereby identifying a fundamental question explored by philosophers since classical antiquity (Quine, 1948). Although Quine’s response “*Everything”* is all-inclusive, lack of explanation of the kinds of things that exist and their interrelationships leaves much room for debate. Investigation of this problem is undertaken within a branch of Metaphysics, namely, Ontology.

Originating from the Latin term *ontologia* which in turn derives from metaphysics (the Greek *logos*, theory and *ont*, being) *ontology* may be translated as the “theory or science of being” (Bunnin & Yu, 2004; Lowe, 2005). The term *ontology* (or ontologia) was coined in the seventeenth century, (Blackburn, 2014; B. Smith, 2004b), specifically stated as 1613 by Smith (ibid.) and more recently stated as 1606 [4]. Yet, examination of the theory of being began as

[4] http://www.ontology.co/jacob-lorhard.htm (Viewed 06/11/2015) The first edition of *Ogdoas Scholastica* by Jacob Lorhard in 1606 contains the first occurrence of the Latin term"*Ontologiae*".

early as the fifth century BC and was the subject of considerable study by Plato and Aristotle (Blackburn, 2014; Bunnin & Yu, 2004; Studtmann, 2011). In modern times major contributors to the field include Immanuel Kant, Martin Heidegger and Willard Van Orman Quine (Bunnin & Yu, 2004; Craig, 1998; Lowe, 2001).

Although general agreement on Ontology as a *science of being* is evident, different approaches and conceptions of ontology have always existed. An outline of the early considerations of two of the most influential philosophers in western philosophy, Plato and Aristotle, follows.

Plato's theory of Forms (or Ideas) concerns the existence of special non-spatial, atemporal entities that are both the objects of knowledge and somehow the cause of whatever transpires in the physical world (Silvermann, 2014). The theory holds that each Form *is* or *has* a unique essence, for example, the Form of Beauty is Beautiful, the Form of Justice is Just and so forth. Unlike Socrates, for whom a Form or universal is inherent in particular things, Plato's Forms are unchanging and independent and the physical world and all therein are a *copy* of the Forms, (Ibid,; Devereux, 2003, pp. 75-99; Irwin, 1995), hence are dependent on Forms. *Partaking* in a Form, by acquiring a Form-copy, together with the effects of material matter *causes* a particular thing to be characterized in some manner, by virtue of *having* something (Ibid.).

In contrast to Plato's Forms Aristotle developed a system of categories. An initial classification is founded on two concepts of beings: "said-of another" and "present-in another" (Studtmann, 2011), from which four classifications of beings derive. These include: accidental universals (said-of and present in); essential universals (said-of and not present in); accidental particulars (not said of and present in) and non-accidental particulars or primary substances (not said-of and not present-in) (ibid). Aristotle's second classification system of things that are said, comprises ten highest kinds or categories: substance; quantity; relatives; quality; somewhere; sometime; being in a position; having; acting and being acted upon (Ibid.; Cohen & Matthews, 1991). Exemplars of primary substances include human being, horse and so forth. B. Smith (1997) provides further clarification on the categories, such that: substances can exist on their own; remain numerically one and the same and can have different properties (accidents) at different times; can stand in causal relations; are one by a process of nature and has no proper parts which are substances themselves; are self-identical while existing continually in space and have no temporal parts. Universals on the other hand, are dependant for their existence on that by or in which they are instantiated, for example, humanity is an essential universal; redness an accidental universal whereas accidental particulars or events may include such things as a smile, bruise, handshake (ibid).

Contemporary philosophers (Klein & Smith, 2010; Quine, 1948) highlight what Quine refers to as the controversy of *universals*, distinguishing three long-standing mediaeval points of view:

Translation of Book 8 is available at https://www.illc.uva.nl/Research/Publications/Reports/X-2008-04.text.pdf (Viewed 06/11/2015)

Realism, referring to the Platonic doctrine of Forms such that universals or abstract entities exist independently of the mind; Conceptualism, accepting the existence of mind-made universals and Nominalism, rejecting universals and abstract entities altogether.

Philosophers such as Lowe (2001) assert that ontology divides into an a priori part that “seeks to establish what categories of entities could exist and co-exist making up a single possible world and an empirical part that seeks to establish what kinds of things do exist in this actual world, based on empirical evidence”. Similarly, Jacquette (2002) distinguishes between Ontology as pure philosophical ontology whose concern is what it means to be or to exist and the putative existents in the world and applied scientific philosophy ‘an ontology’ concerned with elucidating possible categories and kinds of beings for a domain. B. Smith (2004b, p. 155) definition of ontology is “the science of what is, of the kinds and structures of objects, properties, events, processes and relations in every area of reality…. seeking to provide a definitive and exhaustive classification of entities in all spheres of being”. B. Smith (2004b) further distinguishes four schools of philosophy by the diversity of approaches to the classification of entities and relations. These include, *Substantialists* V *Fluxists; Adequatists* V *Reductionists*. Substantialists conceive ontology as *substance* (*continuants)* whereas Fluxists view ontology in terms of *events* or *processes* (*occurants*). Adequatists seek a taxonomy of all entities in reality at all levels of aggregation and Reductionists seek to establish the fundamental particulars, properties and relations which in turn determine all other facts in the physical world.

Rather than enumerating the kinds of things that may exist Quine claims that science and not philosophy serves best to determine what exists (Audi, 1999; Quine, 1948). In answer to the question of ‘What is there’ Quine posits that ‘to be is to be the value of a bound variable’ (ibid), such that, for a formal theory to be true, the variables within the scope of the existential quantifiers (∃) constituting the theory, represent those things that hold or must exist. The term ‘ontological commitment’ refers in a general sense to the set of objects a person takes to exist and thereby constitutes the world and as fashioned by Quine it refers to the objects common to *an* ontology fulfilling a formal theory (Audi, 1999; Bunnin & Yu, 2004; Hookway, 2010). Consequently, the notion of *an* ontology now stands: a theory embodying only a *partial* ontology, committing to the existence of its’ collection of preferred types of objects (B. Smith, 2004b). Lowe (2001, p. 2) however reminds us that “Reality is one and truth is indivisible” , such that for individual theories to be true, they must all fit together and be true of reality as a whole and that determining consistency among partial representations is best left to philosophical Ontology.

2.1.3. Ontology and Computer Science

The term *Ontology* explored in the previous section from a philosophical perspective has been borrowed by computing. In 1967, in the context of data processing Mealy (1967) asserts that the issue is one of ontology, when arguing towards a *theory of data*. In 1980 in the artificial intelligence community (J. McCarthy, 1980) used the term when proposing a commonsense approach to representation. In computer science the term Ontology has been adapted to *an ontology,* denoting a knowledge object or computational artifact (Guarino, Oberle, & Staab, 2009; Uschold & Gruninger, 1996) and signifies the limitation in scope to knowledge of things that exist relevant to a specific theory or domain of discourse. This echoes Quine's view on the role of analytic ontology, outlined in the previous section.

During the knowledge sharing and reuse initiative in the 1990's the knowledge engineering community explored the use of formal ontological principles for content specification of the *real* world: a shift in focus from traditional AI concerns of *reasoning* over what can be represented (Gruber, 1993, 1995; Guarino, 1995). Arising from these investigations, a number of definitions of ontology are posited. Refinements of Gruber's early and ambiguous definition "*an ontology is an explicit specification of a conceptualization*" (Ibid) exist in the literature (Daconta, Obrst, & Smith, 2003; Guarino et al., 2009; Klein & Smith, 2010; Swartout & Tate, 1999; Uschold & Gruninger, 1996).

In the twenty first century ontologies begin to play a prominent role in some domains in computer and information science, the impetus no doubt being the emergence of new technologies and the Semantic Web (Berners-Lee et al., 2001). With increased possibilities for knowledge sharing, reuse and interoperability, the importance of an interdisciplinary approach to content identification and specification of the reality of the domain of discourse cannot be underestimated (Boem et al., 2013). In short, good ontologies are *"reality representations"* (B. Smith, 2004a) (italics specified in the referenced paper). In this context and building upon previous definitions, this author reformulates the definition of a computational ontology as:

A computational ontology is an artifact describing the reality of a domain of discourse or theory, in terms of its structure, classes (types), attributes (properties) and their interrelationships, from a shared perspective, using an agreed vocabulary, explicitly expressed in machine-readable (formal) language.

Structure and components of an ontology: The characteristics of an ontology in computer science (Table 2.1), are described based on sources from academic literature (Antoniou, Groth, Harmelen, & Hoekstra, 2012; Daconta et al., 2003; Huhns & Singh, 1997; Noy & McGuinness, 2001) along with Prótégé ontology builder software.

Classes (concepts, types) are the core components of a computational ontology. Classes represent the categories or types of things of interest in the universe of discourse. Classes are organized into a hierarchical structure, typically a taxonomy, where the root or top class of the hierarchy is generally called 'thing' or 'entity' which subsumes all classes in the structure. Sibling classes defined as direct subclasses of the root represent major divisions of the domain.

A subclass may be defined on a class. A class C is a subclass of another class C^1 if every object in C is also included in C^1. Classes operate on the basis of generalization and inheritance, such that a subclass inherits all of the properties (if so defined) of the superclass and in turn a subclass may have additional relations and attributes defined on it. A subclass is subsumed by its superclass and the superclass subsumes all subclasses defined on. According to Noy and McGuinness (2001) a class may also be a subclass of more than one superclass, thereby being subsumed by both superclasses and inheriting properties from both superclasses. Class names reflect the shared vocabulary for the ontology. Alternatively, anonymous classes may be derived from an expression such as, union, intersection or restrictions on classes.

Relationships (or relations) represent the relationships between classes in the ontology. Besides the subclass relationships defined in the hierarchy other class relationships may be defined, by virtue of a property. A property is a binary relation between classes: the domain and range of the property in this case are both classes. These relations constrain the relations between individuals in the underlying knowledge base by fixing the domain and range.

Attributes (or properties) provide more information about the internal structure of classes; the relationships to other classes and individuals; the properties of individuals and properties of parts (if these are defined). The domain of a property can be a class or instance and the value of a property (Wilson et al.) can be a class, a literal, a Boolean value, a particular data type, restricted to an enumerated list of values and cardinality of the property may also be defined. Properties may possess logical capabilities such as being transitive, symmetric, inverse and functional.

Instances (individuals or particulars) are the particular entities in the domain of discourse. Instances instantiate the class on which they are defined and if the instantiated class is a subclass the instances also indirectly instantiate the superclass. The boundary between an ontology and knowledge base is not always clear however (Hepp, 2008). Opinions vary between those who propose that instances are included in an ontology (Abiteboul, Manolescu, Rigaux, Rousset, & Senellart, 2011, p. 146; Daconta et al., 2003, pp. 186-187; Noy & McGuinness, 2001) and those who contend that ontology serves to represent reality using a shared vocabulary whereas instances or facts should be defined in a separate knowledge base, to answer questions about the shared reality (Fonseca & Martin, 2007; Swartout & Tate, 1999; Uschold & Gruninger, 1996, p. 18).

Table 2.1 Components of a computational ontology

Component	Definition	Example Domain Cultural Heritage
Class	A class represents a category of entities or concepts in the domain of discourse. Classes are organised in a taxonomic (subclass-superclass) hierarchy, where members of a class may share common characteristics.	Classes: Agent, Artwork, Art Movement, Technique
Subclass / Superclass	A subclass may be defined on a class. It inherits all properties of its superclass (if so defined) and is subsumed by its superclass. A subclass may have additional properties defined on it.	Subclasses of Agent: Person, Corporate Body Subclasses of Artwork: Painting, Sculpture, Lithograph
Anonymous or complex Class	Classes can be derived from existing classes for example, through Boolean operators of Union, Intersection and Negation. The union of Oil (a subclass of ArtTechnique) and Painting (a subclass of Artwork) may represent a class 'oil paintings' as opposed to acrylic or watercolour or pastel paintings.	Oil *Union* Painting
Relationships or relations *Property* *Domain* *Range (value)*	Apart from the subclass/ superclass relationship other relationships may exist between classes. Relationships are defined by specifying a property on one class (the Domain) and specifying the corresponding class containing all possible values of the property (the Range):	Property: is_creator_of Domain : Artist Range: Artwork Property: is_benefactor_to Domain: Agent Range: Person
Attribute	Attributes (properties) define characteristics of classes and of individuals. The domain of a property can be a class, or an instance The value of an attribute can be a class, literal, a Boolean value, a particular data type, restricted to an enumerated list of values, and the cardinality of the property may also be defined.	Class: Painting Property: has_width_in_cm Data type: number Instance: "Men of Destiny" Property: has_width_in_cm Value: 61 Class: Painting Property: Genre, Value restricted to one of [History, Landscape, Genre, Portrait, Still Life, Abstract]
Instances (individuals or particulars)	Instances are the particular things of interest, including concrete objects: a painting, an artist, a place, or abstract entities: a time-span or art movement. Instances instantiate the class on which they are defined and (indirectly) the superclass. Attributes (properties) define the facts about the instances.	Property: Title, value "Draughts" Property: has_signature, value "Value: Jack B Yeats"

Vocabulary and **Axioms:** The terms used to describe the components of an ontology form an agreed vocabulary of the domain. These human readable text terms along with informal definitions as annotations or comments ensure shared consistent semantics among stakeholders (Huhns & Singh, 1997; Uschold & Gruninger, 1996).

Table 2.2 Axioms defining the semantics of the domain, expressed in first order logic

Axiom	Meaning
$\forall x$ (Painting(x) $\Rightarrow$Artwork(x))	Class Painting is a subclass of class Artwork.
$\forall x$ (Artist(x) $\Rightarrow \neg$ nonArtist(x))	Class Artist is disjoint to class nonArtist A sub class or an instance of Artist is distinct from or cannot be an instance of both the Non Artist class.
$\forall x$ (possibleMasqueradeExhibit(x) $\leftrightarrow$ Painting(x) $\wedge$ $\exists$(y)(has_theme(x,y) $\wedge$ circus_theme(y))	This defines an anonymous restricted class of potential paintings for selection by a curator for the Masquerade and Spectacle exhibition. This would include all paintings (instances) with a circus theme whereas in reality only twenty three paintings were displayed in the “Masquerade and Spectacle” exhibition.

In contrast to these natural language expressions the underlying language of an ontology may be expressed in a formal knowledge representation language, based on for example first-order logic where *axioms* define the precise semantics (Table 2.2). Semantics in this context refers to a mathematical or logical representation enabling processing or reasoning over the ontology as opposed to normal understanding of semantics by stakeholders (B. Smith, 2004a).

Classification by Ontology Levels: Different levels of ontologies are identified in the literature. These include Top Level, Domain and Task Level and Low Level ontologies.

Top or Upper Level Ontologies specify generic categories such as time, space, measure, quantity, event and so forth (Guarino, 1998; Henriksson, Pradel, Zschaler, & Pan, 2008; Kless, Milton, & Kazmierczak, 2012; B. Smith, 2004b; Sun, Liu, & Li, 2012) and are domain independent. Top level ontologies provide common neutral specifications with potential for base ontology integration and or re-use by communities of users spanning a variety of domain specific ontologies (ibid).

Base or *Domain and task ontologies* represent the reality of a particular domain (for example, art, geography, medicine, aviation) or of a generic task or activity (like diagnosing, selling, purchasing), by specializing the terms in a top-level ontology (Guarino, 1998; Sun et al., 2012).

Low-level Application (or Role) ontologies may span different base ontologies but reside below the base ontology, providing further specialization such that classes often correspond to roles played by domain entities while performing a certain activity (Guarino, 1998; Henriksson et al., 2008).

Other classifications of ontology are also proposed in the literature, for example, Garbacz and Trypuz (2013) identify five orthogonal ways of classifying ontologies, with respect to: Ontological type (either a logical theory or a conceptualization); the domain; formal type of language; complexity of structure and the intended usage of the ontology.

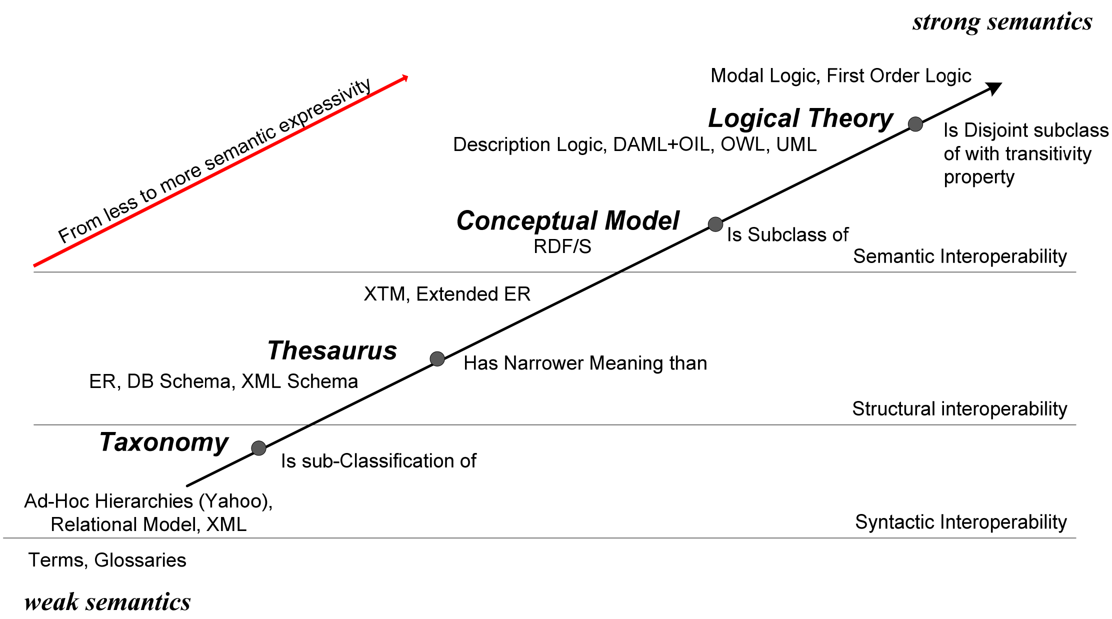


Figure 2.1: Ontology Spectrum – semantic expressivity of resources, based on (Daconta et al., 2003) and adjusted to include weaker semantic resources from (Lassila & McGuinness, 2001) modification of which is published in (Guarino et al., 2009)

Classification by semantic expressivity: The description of the cultural heritage ontology above comprised features of ontology as a formal theory. However, different perspectives on what constitutes an ontology prevail (Boem et al., 2013; Grabar, Hamon, & Bodenreider, 2012; Hepp, 2008) where perhaps usage of a taxonomy and of similar languages for the construction of knowledge artifacts may contribute to the inconsistent interpretation and application of the term ‘an ontology’ across different semantic resources. According to Lassila and McGuinness (2001) three distinguishing features of an ontology include: a) finite controlled (extensible) vocabulary, b) unambiguous interpretation of classes and term relationships and c) strict hierarchical subclass relations between classes. A continuum of semantic representation (Figure 2.1) depicts the weakest semantics at the bottom left to the most rigorous or formal representation with strong semantics at the top right hand side.

Controlled vocabularies such as *terms* (a finite list of linguistic terms describing a concept), *glossaries* (with human readable description clarifying the meaning of terms) (Grabar et al., 2012; Guarino et al., 2009; Hedden, 2010) are lists with no strict hierarchy of subclass relations defined and are classed as very week semantic resources that cannot be considered as ontologies.

Taxonomies or any strict hierarchical structure of classification of terms or concepts may be ontologies or sometimes mistaken as ontologies (Daconta et al., 2003, pp. 146-151) if in fact they reflect *sub-classification of* rather than the semantically stronger *subclass of* relationship. Lassila and McGuinness (2001) observed an example of such weak sub-classification in Yahoo’s early catalogue of services where it seems *dress was a subclass of woman*.

Thesauri and ontologies may also seem similar especially given the taxonomic structure, but are in fact fundamentally different things (Kless, Milton, Kazmierczak, & Lindenthal, 2014). Traditionally a thesaurus is defined as a controlled vocabulary of *terms* generally organized in a hierarchical structure supporting broader/narrower relations (Daconta et al., 2003, pp. 159-161; Harpring, 2010b, p. 24; Hedden, 2010). Thesauri were originally designed for indexing of catalogues and search and retrieval and comprise terms and their semantic interrelationships, including: equivalence; homographic, hierarchical and associative (ibid). Even when thesauri are reengineered in an ontology language supporting formal specification and 'terms' are replaced by 'concepts', a thesaurus concept may arbitrarily correspond to a universal, instance or attribute of an ontology; a variety of subclass relations may be defined (broader-narrower, is-a, associated with, and multiple parents can be specified) such that transitivity does not necessarily hold across the different types of concept relationships (Kless et al., 2014).

Moving further up the scale of semantic expressiveness, conceptual models comprise the primary entities, attributes and relationships and possibly rules that associate same (Daconta et al., 2003). Encoding a conceptual model in an appropriate form, for example OWL (Web Ontology Language) or RDFS (Resource Description Framework Schema) data model captures the semantics of the domain. As such, these encoded models are viewed as ontologies, representing a shared understanding of reality by a user community (Boem et al., 2013) where strict subclass relationships and class properties are specified. In addition to RDFS, the facts or instances of classes may be encoded in RDF (Resource Description Framework)

A Logical Theory, constructed in a formal or logic-based knowledge representation language has formal semantics based on some system of logic extending the semantics of RDFS/RDF. The range of semantic formalisms available depends on the species of ontology language selected. The precise semantics of a theory are defined in *axioms*. For example, axioms for class constructors union, intersection and complement, disjoint classes, restricted forms of quantification and so forth (Guarino et al., 2009; Horrocks, 2008) may be available to formally define the ontology over which reasoning systems may be applied to infer implicit knowledge.

It is interesting to note that the relational database (RDB) depicted at the syntactic and structural interoperability levels, but omitted at the higher semantic level (Figure 2.1), disregards the fact that each database is designed to represent a particular domain or some aspects of reality of interest to stakeholders. Moreover as "a logically coherent collection of data with some inherent meaning" (Elmasri & Navathe, 2014), it describes the semantics of the real world in terms of the kinds of things of interest and their relations, and as such is a representation of that world (Winston, 1992, p. 626). For the purposes of this research the relational model, grounded in mathematical Set theory (Codd, 1970), will be adopted (Chapter 3.).

2.1.4. Discussion: Philosophical ontology applied to Computer Science

Ontology, from a philosophical point of view seeks to establish the possible existence of different kinds of things and their interrelationships whereas an ontology in computer science (Strohmaier et al.) is an artifact representing the reality of a domain of discourse, in a structure reflecting the semantics of the domain.

The literature suggests that there is little consideration of ontology in early investigations in artificial intelligence; *reasoning over that which can be represented* being the primary goal (Gruber, 1995; Guarino, 1995). As adapted by computer science 'an ontology' may be viewed as analogous to Quine's theory of analytic ontology, insofar as it represents a formal theory for a domain of discourse. Leading contributors in the field of ontology such as Smith and Guarino are guided by philosophical principles and practices (Guarino & Welty, 2002; B. Smith, 2004b). These authors argue the benefits philosophical foundation can bring to formal theories in CS to achieve a representation of the truth of the reality and correspondence between classes, properties and instances in ontology in CS and the philosophical principles of universals, accidents and instances/particulars is evident. Furthermore, philosophical methods of dialogue, logic and enquiry are also proposed by other contributors in the field such as Noy, McGuinness and Uschold in the analysis and design of ontologies. These authors also define the essential and preferred characteristics of an ontology which are compatible with philosophical principles. As the field of ontology in CS matures, different levels of ontology considering generic categories or kinds common across diverse domains somewhat similar to Aristotle's categories, is evident. Having said that, and given the potential benefits of these general ontologies Fernández-López, Gómez-Pérez, and Suárez-Figueroa (2013) point out that Upper level ontologies are not being used. Additionally, confusion between resources such as terminology, thesaurus and an ontology is also evident (Grabar et al., 2012) and it appears that some resources currently called 'an ontology' may best fit what is referred to as a 'Concept System' (Klein & Smith, 2010).

Finally, while much of the computer science literature demonstrates awareness of the philosophical roots of the term ontology, it is unclear whether or not a particular philosophical school of thought is applied. It could be argued that an ontology reflects a Realist or Conceptualist view, in that it describes an agreed or shared objective reality of the domain. However, to claim that the Fluxist or Substantialist view or some combination thereof, is, or should be adopted, is surely contingent upon the significance of events over substances, or the accommodation of both, relative to the specific domain or theory in question.

2.1.5. OWL: Web Ontology Language

A previous section discussed the origin of ontology and ontology as adopted by computer science (2.1.3) and this section describes OWL, the language used to build ontologies.

OWL is based on a well defined syntax(s) and semantics and all information is explicitly defined, enabling automatic process and integration of information by machines/computers. The components of an ontology comprising classes, subclasses, attributes, instances and their properties were also previously described (Table 2.1) and are referenced (but not described) in this section.

Precursors to OWL include languages such as XOL, SHOE, OML, RDF and RDFS, OIL, DAML+OIL (Gomez-Perez & Corcho, 2002) but lacking expressiveness and the formal semantics required to enable reasoning, OWL, introduced in 2004 by the world wide web consortium (W3C), addressed these limitations.

Of these earlier languages, OIL and DAML-OIL (Fensel, van Harmelen, Horrocks, McGuinness, & Patel-Schneider, 2001) build upon RDF and RDFS and together with Description Logics (DL) (Horrocks, Patel-Schneider, & van Harmelen, 2003) are fundamental to OWL. OWL however, is presented as a family of languages where three species or sub-languages exist: OWL Full; OWL DL and OWL Lite (Table 2.3). These different sub languages fulfill certain aspects of the requirement of "*efficient reasoning support and convenience of expression for a language as powerful as a combination of RDF Schema with a full logic*" (Antoniou & Van Harmelen, 2004). An ontology designer must select the appropriate sub-language depending on the required constraints and reasoning required.

Table 2.3 OWL families

Name	**Constraints**
OWL Full	All language constructors may be used in any combination but must be compatible with RDF
OWL DL	1. A resource may be only ONE of a class, a datatype, a datatype properties, an individual, a data value or part of a built-in vocabulary. For example, a class cannot be at the same time an Individual 2. Each of the above must be stated explicitly 3. Object properties and Datatype properties are disjoint. For example, inverse, functional, inverse functional properties and symmetric characteristics cannot be specified for Datatype properties 4. Cardinality constraints cannot be placed on transitive properties 5. Anonymous classes are only allowed in the domain and range of owl:equivalentClass and owl:disjointWith and in only the range of rdfs:subClassOf
OWL Lite	Must be an OWL DL ontology, but the following constructors are not allowed: owl:oneOf owl:disjointWith owl:unionOf owl:complementOf owl:hasValue

A short description of OWL elements presented (Table 2.4) is based on academic literature (Antoniou & Van Harmelen, 2004; Horrocks et al., 2003; McGuinness & Van Harmelen, 2004) and W3C specifications (W3C, 2012).

Table 2.4 OWL elements

Element	Description
Classes	OWL elements comprise classes and sub classes and many be defined as disjoint and equivalent classes express as owl:Class rdf:ID, rdfs:subClassOf rdf:resource, owl:disjointWith rdf:resource and owl:equivalentClass rdf:resource respectively
Properties	Object properties relate objects for example the property 'isCreatedBy" relates a Painting (domain) to an 'Artist' (range) Datatype properties relate objects to values such that YearOFBirth must be set to an integer
	Inverse properties may also be set such as "isCreatedBy' and 'Creates'
Restrictions	Restriction may be placed on a class, for example, IrishCollection must be created by artists of Irish Nationality owl:allValuesFrom from the class of IrishPainters or create a new anonymous class
Special Properties	TransitiveProperty defines a transitive property such as 'is ancestor of', 'is older than'
	SymmetricProperty defines a symmetric property such as 'is sibling of'
	FunctionalProperty defines a property that has at most one unique value for each object such as 'height', 'width', 'medium'
	InverseFunctionalProperty defines a property for which two different objects cannot have the same value for example 'isthePassportNumberOf', or 'istheAcquisitionNumberof'
	OWL 2 AsymmetricObjectProperty defines a property such that if it connects A with B it never connects B with A, for example, 'hasChild'
	OWL 2 ReflexiveObjectProperty defines a property that relates everything to itself, for example, 'hasRelative. IrreflexiveObjectProperty defines a property that means no individual can be related to itself by such a role. for example, 'parentOf'
	OWL 2 DisjointObjectProperties defines that two properties are disjoint if there are no two individuals that are interlinked by both properties. For example, 'hasParent' and 'hasSpouse' means that parent-child marriages cannot occur:
Boolean Combinations	Union, Intersection and complement of using owl:unionOf, owl:intersectionOf and owl:complementOF clauses to create new or anonymous classes such as a class of Irish painter born in the 18th century
Enumerations	An enumeration is a owl:oneOf element by listing all of its elements
Instances	Instances of classes are declared using RDF for example to define the artist identified by the number 20101, <artist rdf:ID="20101">
Datatypes	These include types such as string, integer, boolean, date and time. Richer datatypes are introduced in OWL 2

The latest update to OWL 2 dates back to 2012 and retains all of the features of OWL, “albeit possibly under different names” and is backward compatible. According to the W3C documentation, additional features and functionality in OWL 2 offering new expressivity, include, keys; property chains; richer datatypes and date ranges; qualified cardinality restrictions, asymmetric, reflexive and disjoint properties and enhanced annotation capabilities.

Furthermore, new profiles OWL 2 EL, QL and RL, suitable for large scale ontologies (EL), conjunctive queries using relational database technology (QL) and RL is suitable for relatively lightweight ontologies organizing large volumes of individuals.

2.1.6. Topic Maps

A topic map (TM) is a standard for knowledge integration and management that facilitates the discovery of related knowledge stored in disparate resources such as databases, documents, web resources and images, for example. The origin of Topic Maps can be traced to the 1990's when the Davenport Group (players in the computer industry) sought to interchange documents between different computer systems by merging the indexes of the documents into one index (Pepper, 2006).

The standard ISO/IEC 13250, was first published in 2000 and comprises three basic concepts a) topics b) occurrences and c) associations and these describe the names, the properties and relationships of subjects (Garshol, 2004; Pepper, 2006; Steiner, Essmayr, & Wagner, 2001). Topic maps are stored as XTM (XML Topic Map files), the Topic Map standard. The concepts of Type, Scope, and Facet enhance the power of the topic map.

A topic represents a subject such as person, event, place and anything of interest in an application or domain. Different names can be assigned to topics representing the same subject (synonyms), and alternatively, different topics can have the same name (homonyms). An occurrence relates the subject (represented by the topic) to an information resource, by specifying the URI of a resource, for example, a painting or a sketch or a letter in the archives or by storing a string directly in the topic map. Associations represent relations between topics.

Types can be defined for topics, occurrences and associations and they describe the world from which the topics are taken and therefore Topic Maps can describe any domain and any kinds of relationships. Types are selected by the creator of the topic map depending on the system and information resources, where for example types such as artists, artworks, correspondences, sketches, time period, places and images may apply to this study.

The type, occurrences, associations and scope all serve to distinguish the meaning of the topic and the context and kinds of relationships between the topics. By way of example, the topic painting may refer to an artwork type or to decoration type, or Liffey may refer to a river or to a painting and would be resolved based on the type, association or occurrence. Since topics can be *typed,* topic maps facilitate more powerful search and discovery, by allowing a user to find only the artwork and to exclude other kinds of subjects of the name.

Moreover, occurrences also have *types* to distinguish between the different kinds of relationships to the resources. By way of example, to distinguish a sketch from a photo or a portrait from a biography and the *scope* may distinguish the context of the topic; as a letter written by an artist or received by an artist or if the materials have been destroyed (Garshol, 2004).

Associations can also be typed and an association role denotes how the subject is involved in the relationship. For example an object in the archives (a letter) may be related to a painting based on an artist name, the creator/role of the sketch, the content or inspiration for a named painting and so forth.

According to Pepper (2006) these concepts and types and scope and facets demonstrate how Topic Maps incorporate and also extend the traditional knowledge structures from indexes, controlled vocabularies and hierarchical structures of taxonomies and thesauri and provide a unified and flexible graph representation for the classification of information. Furthermore, merging arbitrary Topic Maps into a single Topic Map enables high level knowledge enrichment and knowledge sharing over distributed information resources.

However, creating a Topic Map (TM) manually can be an arduous task and associated research into automatic content tagging (Waltinger, Mehler, & Heyer, 2008), automatic TM generation using relational databases (Eslami & Nazami, 2011), applying statistical solutions (Blei, 2012) and generating an RDF/OWL file from text documents using text mining techniques (Garrido et al., 2013) and search techniques (Yuta, Masaomi, & Yuki, 2017), continues.

2.1.7. OAI-PMH: Harvesting metadata from archives

OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting) is a framework for interoperability of federated content, where content providers adopt the OAI protocol to expose their metadata and in turn service providers adopt the protocol to harvest and process the metadata, facilitating search and discovery services across federated content.

The Open Archives Initiative (OAI) emerged in 1999, comprising a body of scholars intent on developing new mechanisms for efficient dissemination of content within the E-print community (Lagoze & Van de Sompel, 2001). This was a time that witnessed the growth of personal computing, networking and web services with the potential to leverage these technologies to provide a viable alternative and an extension of the services beyond the traditional publishing paradigm (Ibid). The E-print community refers to electronic academic or scientific documents such a pre-print and post-prints of research papers, book chapters, results of experiments and other related materials (Pinfield, 2002).

The distinguishing factors in providing impetus to the initiative comprise: delays in dissemination and access to scholarly texts, rapidly expanding activities of digital libraries and author’s intellectual rights (Lagoze & Van de Sompel, 2001; Rusch-Feja, 2002). By 2002, version 2 of the OAI-PMH was widely recognized as a vital component of the digital information infrastructure (Lagoze & Van de Sompel, 2003) and the initiative extended beyond the E-prints community to include publishers and reference libraries, recognising the potential opportunities to develop new services based on the harvested data (Needleman, 2002).

The OAI-PMH framework comprises the following elements:-

a) Repositories
b) Providers: Content Providers and Service Providers
c) Metadata standard
d) Protocols

First, it is necessary to clarify that in the context of OAI-PMH, the term 'archives' refers to *repositories* of content and not to the more sophisticated or advanced systems for curated materials preserved by professional archivists (Lagoze & Van de Sompel, 2001; Pinfield, 2002). Furthermore, the term 'open' in the context of OAI-PMH refers to *interoperability* based on the availability of metadata harvested from content providers, as opposed to free of charge or unlimited access. Content providers comprise academic institutions, reference libraries and publishers of scholarly or scientific materials that expose the content of repositories through an agreed schema. Service providers, who can also be content providers, harvest the metadata from the content providers and provide search and discovery services across the federated repositories. Due to the widening of the community and content, beyond scholarly papers, the OAI selected the already adopted Dublin Core (DC) metadata standard (2.3.1.1) as a means to achieving interoperability while also encouraging participation in this initiative. Moreover, adopting unqualified DC facilitated resource discovery, (albeit facilitating coarse granularity discovery) supporting the underlying motivation for the initiative and the process was further simplified by specifying that all metadata elements were optional. Whilst no restrictions were placed on the number of elements or sets of elements to describe an item in a repository, the protocol restricts the structure for exposing the metadata to XML documents, with an XML schema for validation.

An XML document comprises a) a *Header* containing an identifier or key to locate the particular item in the repository and the date of the item, b) the *metadata* containing the description or value, for each named metadata element and c) an optional *About* item that typically describes access rights or conditions of use. Based on the above definition, content may be harvested by date or date range. To facilitate more selective harvesting, content providers may define a 'set' or 'grouping' of items, enabling extraction of metadata for all items in the group.

The protocol operates within HTTP over the web, using the GET and POST methods and each request contains the URL of the server of the repository followed by keyword arguments in the form of key-value pairs, that at a minimum specifies the name of the request.

Table 2.5 Requests or functions available in OAI-PMH

Request / Verb	Purpose	Keyword Arguments
GetRecord	Retrieve an individual record's metadata from a repository	Identifier or key/Value metadataPrefix/Value
Identify	Retrieve information about a repository	none
ListIdentifier	Retrieve a list of identifiers of records that can be harvested and optionally specify a date/date range or a set	optional: UTCDatetime/Value (from and until) optional setSpec/Value
ListMetadataFormats	Retrieve a list of the metadata formats supported by the repository, or available format for one item	optional: Identifier/Value
ListRecords	Retrieve/harvest records from a repository and optionally specify a date or a set	optional: UTCDatetime/Value (from and until) optional setSpec/Value
ListSets	Retrieves a list of the metadata sets available in the repository	

Six different types of requests (Table 2.5) allow service providers to harvest a record's metadata; to retrieve information about a repository; to retrieve a list of all items that may be harvested; to list the metadata formats of items in a repository; to harvest records and to retrieve a list of the sets available in a repository (Lagoze & Van de Sompel, 2001; Lagoze, Van de Sompel, Nelson, & Warner, 2015; Needleman, 2002). Optional arguments facilitate selective retrieval by Date, by Set and by Metadata type.

Since its introduction, OAI-PMH has been adopted by numerous institutions and provides valuable discovery services for cross-disciplinary research and the Mellon Foundation has provided several grants for research in OAI in the cultural heritage domain. Focusing on harvesting archives encoded in the more complex metadata standard, Encoded Archival Description (EAD, in 2.3.1.3) early research demonstrated some difficulty in mapping EAD to OAI-PMH. Moreover, the research showed that multiple records would be needed to map an archival record (collection) rather than one identifier record and that although it might be consistent enough for the finding aids, more research would be required to output EAD in OAI (Prom & Habing, 2002; Prom, Rishel, Schwartz, & Fox, 2007). Furthermore, and in a broader context, criticisms have surfaced in relation to the loss of more complex data due to the limited dataset of Dublin Core (DC) and the availability of only one unqualified Date element in DC (Corrado, 2018; Devarakonda, Palanisamy, Green, & Wilson, 2011).

2.1.8. Relational Databases

Introduced in the 1970's, the relational database revolutionized the computational representation in large scale data banks (Codd, 1970), with the development of techniques to accurately describe relations (set of entities) and their interrelationships, resulting in reduced structural redundancy and increased data consistency. Moreover, the relational model enabled both logical and physical independence of data, easing the burden on developers of having to know the exact order or position of stored data and how to locate it, thus minimizing amendments to application programs on changes to the internal data representation and physical independence minimized disruption and rework for database access from a variety of different machines and storage formats.

A database is an organized collection of data, about a particular domain, or process or some aspect of reality, referred to as the Universe of Discourse (UoD) or miniworld (Elmasri & Navathe, 2014, pp. 2-3). Furthermore, since it is designed for a specific purpose, a database contains a coherent collection of related data, with some inherent meaning, shared by the user community. According to Codd (1979) a relational database is a time-varying collection of data, presented in tabular form, comprising relations of assorted attribute types and of varying cardinality. Currently, relational databases underpin countless systems worldwide, including commercial, governmental, banking, medical, reservation systems, enterprise resource planning and customer relationship management systems, to name but a few. Common characteristics of these systems include, large scale shared data banks, multiple user access, multiuser transaction processing, backup and security and access controls.

General purpose software systems to create and maintain databases, known as relational database management systems (RDBMS), incorporate a toolset to define, organize, maintain and manipulate databases. Furthermore, the RDBMS maintains a catalogue of all metadata about the structures (schema) and constraints defined for each particular database and this catalogue restricts access to the database and operations (to authorized persons only) and controls database operations, safeguarding database integrity.

In the relational model, relations (commonly referred to as tables) contain all of the data, where each relation represents a specific entity type or class (for example, person, artwork and exhibition) in the UoD. Codd (1979) describes the properties of relations as follows. Instances or individuals contained in tuples (commonly referred to as rows) in relations, uniquely identified by a Primary key. Attributes represent atomic values describing characteristics of an entity, and the type or domain of the attribute constrains the value stored for the attribute. The order of attributes within a tuple is insignificant and the order of tuples within a relation is also insignificant. Relationships between entity types are enforced through a foreign key value in one tuple, referring back to a primary key in another tuple. Tables, columns in a table, and relationships are assigned unique names, all of which are recorded in the RDBMS catalogue.

Similarities between the relational database (RDB) model and the ontology model (described in Ontology and Computer Science) are apparent. Each particular database is a computational artifact representing aspects of the reality of a domain, in a structure comprised of entity types (types or classes), attributes (describing the characteristics of individual entities) and relationships between the entities. Constraints may also restrict the content or values allowed in the database. Furthermore, 'Views' may be defined to represent subsets of data (by way of example, all oil paintings in the database, all landscape paintings by a particular artist, all sculptures created within a particular time-span etc.). Alternatively, 'views' may be defined to combine entity types (by way of example, combine the artist details with all or some of the details of a particular work of art, including a description of the style/movement, the methods and techniques used to create the work). Moreover, since all data in a database are stored in base tables, views are generally only materialized (populated with data) as a result of a query on the database. Finally, employing metadata and vocabulary standards developed for the domain, in the design of the database schema, facilitates consistency in organization and description and supports interoperability.

Unlike the computational ontology model, however the meaning of the database is not always explicitly stated, rather, it is evidenced through the organization and combination of data elements in tables that represent entity types and on which relationships are defined between the entities types. Therefore, it may be more difficult for the user community to understand the data structures and terminology used in database technology than that used in a computational ontology.

Relational database technologies underpin the majority of vendor supplied museum systems (Table 1.1), while some use Object Database systems, NoSQL systems and other employ proprietary storage systems. Employing a relational database (RDB) is well suited to describe the complexities of the domain in the computational representation for this study, which incorporates the knowledge of domain experts. The complexity of aesthetic objects, these temporal and spatial entities, arises in part, due to the many and diverse relations between the object and other entities. These entities include, for example, the artist; sitter; theme; subject matter; description; captions; the materials constituting the work of art, the image or images of the work, provenance of the work of art; critical reviews; relationships between an artwork and other artworks; and especially relevant to this study, the relationships between the artwork and archival materials. Moreover, RDB technology is also well suited to capture the range of 'transactions' or 'events' typically associated with a work, including acquisition, conservation, exhibition, provenance, loans out / in and so forth.

Today, the relational database is but one of a number of extant storage models, including more recent technologies, such as NoSQL, triple stores and linked data. Although introduced almost fifty years ago, continued global usage highlights the significance of the relational model (Coronel & Morris, 2016, p. 43; Van Hooland & Verborgh, 2014, p. 14) such that Van Hooland et al assert that "*relational databases are here to stay, and will not be disposed of in favour of triple stores*".

2.2. Works of art and cultural heritage

Designing a computational representation of works of art requires a clearer understanding of the various classes in which collections of works of art can be organised and the relationships that exist between works of art and between works of art and their collaterals, and these aspects are explored in this section.

Considerable scholarly discourse about art, from art historical; philosophical; aesthetic; psychological and visual culture perspectives, exists. Yet, a precise, universal definition of art cannot be formulated, due to the complex nature of art and the diversity of artworks, with terms such as *representation*, *aboutness*, *embodiment* and *aesthetic* ascribed to works of art in the literature. The ontology of art, the philosophical inquiry into the nature of works of art, explores these complex objects (2.2.1). A review of art specialist concepts and practices provides further insight into the intentional and cultural relations of art (2.2.2). Existing case studies on art research consider the needs and resources used art specialists when researching a work of art (2.2.3) and the section concludes with an exploration of primary sources, artist archives, that may contain rich resources enabling deeper insight into a work of art (2.2.4).

2.2.1. Ontology of Art

The philosophy of art and its relationship to man and beauty has engaged thinkers such as Plato and Aristotle through to Kant, Dewey and Heidegger among others in both the classical and modern eras (Hofstadter & Kuhns, 2009). According to (Lamarque, 2012, p. xii) innovative research into the ontology of art by philosophers, including Arthur Danto; Nelson Goodman, Richard Wollheim, Joseph Margolis; Amie Thomasson; Gregoiy Currie; Eddy Zemach; Jerrold Levinson among others, has resulted in a much clearer understanding of how works of art fit into the general scheme of things.

The ontology of art attempts to understand the ontological status of works of art such as painting; sculpture; musical works; literature; drama; prints and other art forms. Determining what *kinds* of artworks exist and what kind a particular artwork is, requires examination of these different art kinds with respect to their *existence, identity* and *persistence* conditions. (Margolis, 1977; Margolis, Currie, Martin, & Van Camp; Thomasson, 2004; Walton, 1970). This task however is more complex than it seems, given the diverse perspectives about works of art proposed in the literature (Table 2.6).

Table 2.6 Diverse opinions on works of art as surveyed in the literature by Thomasson, 2004; 2005.

(Thomasson, 2004)	"a denial that a work of art is a physical object because of the creative, imaginative process of the artist and the imagination required by the viewer to experience it"
	"a work of music may be composed and exist only in the 'head of the artist"
	"works of art are unreal objects, because seeing the aesthetic object requires imaginative acts of consciousness."
(Thomasson, 2005)	"all works of art are action-types rather than individuals, including painting and non-cast sculpture"
	"Literary works are abstracta that cannot be created or destroyed"
	"plays not only present fictional characters and events but are themselves fictional objects"
	"transcription of a musical work must be a different work"
	"no painting or sculpture can be repaired or restored"

If these opinions on works of art hold, then Thomasson (2004) surmises that works of art are abstract objects and cannot be destroyed so long as they remain under consideration as aesthetic objects. These opinions may complicate the task of determining the ontological status of art works but perhaps the question is, which if any of these views might be acceptable? A more practical view (Margolis, 1977) claims that particular works of art cannot but exist as embodied in physical objects and being cultural objects they possess intentional properties such as design, expressiveness, symbolism, representation, meaning, style and so forth.

Considering the notion of *singular* and *multiple* artworks (Levinson; Weh, 2010) it is plausible that *cast* works and *prints* are multiple or a finite number (Ibid.) Weh asserts that there can be only one exemplar for a singular artwork, for example, there is no type that the painting Mona Lisa would be a token of, except the original painting; the original *is* the work. Moreover, copies and reproductions are just that. On the other hand, even if only one instance (a manuscript or a score) of a novel or symphony existed, the fact is that there *could be* many instances, such as publications, recordings and performances (Margolis et al.; Thomasson, 2004; Weh, 2010). Paintings as *singular* artworks reflect both social and art specialist practices. Questioning whether practices may be misleading Weh (2010) cites Nelson Goodman, who claims that *a)* a copy does not offer the same aesthetic experience, *b)* there is no notation and *c)* originals and their copies differ in their history of production. Whether one can discern a copy or forgery is arguable [5] (Lamarque, 2012, p. 149) and aesthetic experience is dependent upon that. Sketches or preliminary drawings are not notations for sculptures or paintings in that the artist does not include instructions on how to accomplish these works and as Gombrich (1995, pp. 34-35) notes a sketch may not even include the main expressive features of the finished artwork. If only *c)* holds, then differentiation between singular and multiple depends on the artist's intentions, without knowledge of which, common cultural practice of artists and society of that period must prevail (Thomasson, 2005; Weh, 2010).

Determining the ontological status of an artwork extends beyond the physical, functional or resemblance characteristics, to consideration of the artist's intention to create a work of *a kind,* and art specialists or 'grounders' are necessarily involved in the task (Davies, 2009; Thomasson, 2005). Art specialists bring their background concepts and knowledge, embedded in their practices, to the task of determining what ontological conditions of the work are upheld (Ibid.). Contrary to Thomasson's 'simple method' Dodd (2012) defends a 'discovery method', suggesting that by starting with a proposal such that X is an F (for example, this object is a Painting) and that because the facts are mind-independent, the ontological investigation may in fact reveal mistaken assumptions about the ontological nature of art-kinds. Dodd defends this method because it, he claims, demands explanation, whereas "in the art-specialist world the constraint of explanation is missing." (Ibid).

[5] Eight Of The Biggest Art Forgeries of all Time http://www.businessinsider.com/art-forgeries-2011-6?op=1&IR=T

The *existence* of works of art depends on a number of criteria. Setting aside the matter of unfinished works by famous artists, the first condition must be that the artist has completed the task: that the artist has finished working with the medium (canvas, paint, stone, bronze, words, sounds, patterns of movements) of the object and has created an intentional artwork (Lamarque, 2012, pp. 33-35; Thomasson, 2005). Secondly, the art specialist community approves the artwork as possessing essential intentional, artistic and relational properties (Ibid.). Furthermore, established art specialist beliefs and practices not only judge and herald a new artwork, but also constrain the *persistence* of the artwork. Notwithstanding the fact that works may be destroyed or lost or decay beyond repair "works cannot survive as artworks if art specialist practices are lost" (Lamarque, 2012, p. 54) . Concerning the *identity* of works Danto (2001, pp. 33-36) highlights the fact that any two or more *different* artworks of the same kind may indeed appear to be the same, ascribing to these the term "indiscernibles". Thus, individuating artworks (whether indiscernables or forgeries) cannot rely solely upon intrinsic perceptible properties, but must extend to contextual properties and relations such as who, what, where, why and when, because these properties also characterize the work (Ibid.;Lessing, 1965).

Conceptual art and new or emerging art forms bring with them new challenges to social and art specialist practices. Responding to new art kinds such as new medium, physically damaged or decaying artworks, artworks in which the museum visitor partakes or participates, requires adjustment of ontological investigation (Thomasson, 2010). The notion of no notation or instruction previously noted may be overthrown in some cases (Lamarque, 2012, p. 38; Thomasson, 2010). Investigations may now include the '*artist sanction'*: a public statement about the conditions in which the artwork should be displayed, instructions to curators about its display and how it is to be preserved, which features are essential and so forth, substantially amounting to a specification of the ontological status of the work (Ibid.).

In conclusion, perhaps much of the problem of the ontological status of artworks reflects the nontrivial characteristics of works of art as aesthetic and cultural objects with complex intentional relational properties, as opposed to ordinary functional objects in the world (Lamarque, 2012, p. 56; Ibid.). Notwithstanding the diverse viewpoints reported above, and concurring with Margolis, Lamarque asserts that "works of art are real entities, they can be seen, heard or touched, as appropriate; they are the result of a creative and imaginative process by the artist, to be observed and experienced by the public".

2.2.2. Art Specialists perspective and practices

The previous section highlighted the primacy of art specialist theories and practices in relation to the ontological status of artworks. Examination of art specialist practices concerning the appreciation, criticism, evaluation and judgment of aesthetic objects however reveals two opposing theses, with a variety of formulations proposed in between. Aesthetic Empiricism or Formalist theory proposes that detecting all relevant aesthetic value to appreciate and judge a painting requires merely looking at it along with the immediate experiential encounter (Bell, 1913, pp. 1-37; Lessing, 1965). On the other hand, the anti-empiricist or *contextualist* theory rejects empiricism on the grounds that crucial properties and relations necessary for artistic appreciation and evaluation should include those properties and relations beyond or not manifest in the artwork (Danto, 2001, pp. 33-36; Davies, 2009; Margolis, 1979; Walton, 1970).

Advocates of the empiricist thesis focus on form and aesthetics over representation and originality of artworks. For example, art critic Clive Bell (Ibid.) maintains that art appreciation requires only a clear mind and must start as a personal experience along with knowledge of 'Significant Form'. Significant form, a quality common to all visual artworks, involves the combination and arrangement of "line and colour, certain forms and relations of forms" and knowledge of three-dimensional space (Ibid.). Similarly, (Lessing, 1965) claims that an aesthetic experience is "wholly autonomous"; entities and facts *about* the artwork are irrelevant. Influences of the principles and methods developed by art historian Heinrich Wölfflin with respect to Form are apparent in this approach (Wölfflin, 1932), but Wölfflin also considered comparison of artworks from different styles with respect to form and subject essential to the practice of art history. The empiricist approach is reminiscent of Russell's knowledge by acquaintance and it assumes a degree of artistic sensibility on the part of the viewer. Indeed, considerable research by art specialists, artists, art critics, art historians and curators supports the development of this sensibility (Berger, 2008; Dondis, 1973; Hodge, 2014; Rorty, 2014), where parallels between language and visual literacy are evident. Visual literacy, however, is complex and extends beyond just seeing, beyond subjective preferences towards an understanding of visual concepts and methods of seeing and a universality of meaning and communication (Ibid.;Dondis, 1973, p. 182). Undoubtedly, publicity recognizes this communicative power of art, often using "images of paintings to lend allure or authority to their own message…suggest a cultural authority, a form of dignity, even of wisdom " (Berger, 2008, pp. 123-143).

Understanding human relations with images has been the subject of empirical research conducted over decades, some of which distinguishes for example, viewing patterns, emotional and cognitive differences between art expert and novice reaction to images. A recent small study measuring eye movement and fixation of six art experts and eight novices viewing twenty abstract paintings revealed that art expert knowledge and sensitivity to formal concepts override stimulus-driven factors (such as colour and luminance) in attending instead to visual patterns whereas the novices fixated on the more stimulus-driven salient features (Koide, Kubo, Nishida, Shibata, & Ikeda, 2015). Another recent study measuring physiological indicators (facial electromyography) of emotional response, along with self report, from eighteen low-, eighteen middle- and twenty high-art expertise participants, viewing 64 artworks and forty IAPS

[6] pictures (equally split between negative and positive valence) revealed that art experts show aesthetic and emotional processing differences for visual stimuli with attenuated reactions and less extreme valence rating compared to the non-experts (Leder, Gerger, Brieber, & Schwarz, 2014). In this study the art experts also liked negative art more.

From the art historical perspective, art historian Ernst H Gombrich reminds us of the ambiguity of the visual image, the need to decipher the image, that awareness of form (for example, figure and ground, shape and direction, foreshortening and perspective) assists in revealing the real problem posed in the image. Moreover, understanding the authentic nature of an artwork requires additional knowledge such as the intentional and cultural relations influencing and reflected in the artwork (Gombrich, 1995). The contextualist thesis, supporting this view, underpins the prevailing theories and practices of art specialists (de Rynck, 2004; Hodge, 2014; Thompson, 2006) such that social and cultural history and conventions, knowledge of form, technique, other artworks and the artist biography may be considered, as appropriate to assist in understanding paintings.

Art historian Erwin Panofsky formulated a methodology for discovery of the true or "intrinsic meaning" of an artwork, as a reflection of the cultural context in which it was created. The methods of 'Iconology' focus on three-levels of meaning (Panofsky, 1972, pp. 3-16), where the objects of interpretation comprise: Primary or Natural Subject- Level 1; Secondary or Conventional Subjects – Level 2 and Intrinsic Meaning or Content – Level 3. The results of Level 1 represent "the world of artistic motifs… enumeration of the content and form providing a pre-iconographical description of the artwork", derived through a pseudo-formal analysis of artistic elements (such as line, colour, shape, texture and so forth) and *natural* objects, (for example, shapes of people, animals or other objects). Eliciting *expressional* qualities requires analysis of the combination and arrangement of the elements and objects depicted. (Ibid.) In level 2 (Iconographical analysis- in a narrow sense) the viewer determines the subject or theme of the artwork through analysis of the pre-iconographical description in conjunction with knowledge about the objects depicted and of conventions of artistic representation of the particular objects or figures, in the period in which the artwork was created. Panofsky refers to this as "the world of specific themes or concepts manifested in images, stories and allegories" (Ibid. pp. 6). In the final level (Iconographical Interpretation–in a deeper sense), the viewer considers the artwork in the broader historical and cultural context and how prevailing attitudes and traditions influence the artist (unconsciously) and are reflected in the artwork. (Ibid. pp 7-8). It is clear that considerable research of texts and other artworks may be required at all stages of the methodology, in order to recognize objects beyond the knowledge of the viewer and to

[6] International Affective Picture System, a set of affective/emotional images used in psychological research

familiarize oneself with other artworks and representational conventions and symbols of the period and the culture in which the artwork is created.

Particularly relevant to professional experts, such as museum curators, however, is the need for art historical practices of provenance researchers, to engage with ethical acquisitions policies (V. Reed, 2013) rather than relying on the dealer or donor to provide provenance documentation. In the art world, provenance is defined as “the history of the ownership of a work of art or an antique, used as a guide to authenticity or quality; a documented record of this” (OED), “the chain of ownership” (Feigenbaum & Reist, 2012, pp. 48-49). During much of the twentieth century museum staff may have ignored warning signs of associations with Nazi looting of World War II and the author suggests that implementing strict acquisition policies and conducting and documenting their own provenance research may avoid future compensation claims or restitution of objects, to the rightful owners. Further emphasizing the necessity for institutions to carry out their own provenance research, to prevent cases of fraud and forgery, Sandis (2016) cites notable cases detected as recently as May 2000. Acquisition guidelines by the American Alliance of Museums and the Association of Art Museum Directors (V. Reed, 2013), international regulations “1970 UNESCO Convention”, International Council of Museums Code of Ethics 1986 and Swiss legislation requiring due diligence (Levine, 2009) should preclude unethical acquisitions in the future.

Feigenbaum and Reist (2012, pp. 17-18) demonstrate that provenance details may in fact appear on the physical artwork itself, noting that major collectors frequently mark the back of old-master painting with red wax seals of the owners name and that typically marks are put on the paintings when they enter royal collections. Additional mechanisms for tracking provenance include mounting paintings in custom designed, recognizable frames and placing inventory numbers on the front a painting. In contrast to back markings on Western art, a Chinese scroll can present a wealth of provenance information (embellished with colophons and seals of the owners) on its front surface. Moreover, these markings are welcome in China, especially from a collector’s perspective; they often supplement and enrich the viewer’s experience of a painting (Ibid pp. 32 - 35). Alternatively an imperial owner could establish a connection to the original painter of a handscroll through the shared interest in the calligraphic practice of their poems (Ibid p. 38).

2.2.3. Art Research: methods and needs of domain specialists

Implementing technologies of benefit to the art expert, scholar and professionals conducting research in the arts, requires an understanding of the nature of research in the arts, including the methods, types of research and information needs of these entities. This understanding may then inform the design of technologies in support of art research and will ultimately determine correspondence between technologies and specialist user’s requirements. Literature and case studies reviewed below provide some insight into this matter.

A review of the literature on art research processes and needs (Beaudoin, 2005) traces the "relatively slow and hesitant adoption of technology" by art historians, whose methods involve searching print sources of monograph; bibliography; general reference materials; database indexes; slides, digital images; photographs and consultations with colleagues. Descriptions of the art historian's research process places the art object itself at the centre of the process with visual memory recall being augmented by comparison of an abundance of images: comparison being "the methodological basis of the discipline of art history" (Brilliant, 1988). According to the author, establishing the artistic, social and historical relationships necessary for interpretation of the object requires access to catalogues, standard reference works, bibliographic materials, older books and research articles: the majority of which at that time were mainly housed in physical repositories, apart from online resources from Columbia and Stanford universities.

An early case study into the working methods and information needs of college-based and museum-based art historians (Stam, 1997) characterizes the process as a "fairly private activity", a "contemplative undertaking involving objects of art". Similar to Brilliant, the study affirms the art object as the central and primary information source, and that respected opinion and information on other related objects is used extensively to assist with interpretation of the object. The forty nine participants reported undertaking extensive use of in-house and external libraries (including building personal libraries of books and images); reading journals; communication with colleagues and subject experts referred to as 'the Invisible College or Reference Group' and 'The Gatekeeper' along with travel to museums, to complete their research.

A small case study of fifteen art historians (Rose, 2002) found technology being used in the initial information gathering phase of research, alleviating some problems of time, travel and access to images. However, the Library was used by 100% of participants with Archives coming in second at 27% and the Internet at only 15%. Importantly for those working to support research in the arts, including information technologists, the study distinguishes the methodological and theoretical approaches of art historians listing items such as connoisseurship; iconography; cultural studies; gender issues; materials and construction; biographical; scientific analysis; social; political; theory and criticism among many other areas of concern to art historians. Given the intellectual and interdisciplinary nature of the research it is not surprising then to find that the participant's wish list included, among other items: more access to intellectual content such as scholarly journals with images, search by more complex subject terms and improved access to Eastern Asia museums.

The complexity of the problem of intellectual access to visual objects arises in part due to the 'multiplicity of concepts' associated with an art object (Beaudoin, 2007), such that subject indexes of an artwork may be at the level *of* (whatever is visible in the artwork) and also at the level *about* (which may be something entirely different than what is depicted in the artwork). The author notes that the time required to document subject at these levels may be prohibitive, but that standardization of metadata and thesauri may go some way towards solving the problem. Additionally, the varying information needs of the different user groups highlights the volume and range of information required. By way of example, museum registrars are concerned with the condition; location; logistics; curators and scholars may be concerned with provenance; exhibition history; subject terms and so forth (Ibid.), whereas the general public might be satisfied with standard information and a description of the art object.

Considering the case of art historians and the supposition of their reticence to using technology Gaehtgens (2013) informs of the changing landscape, noting that many institutions now employ art historical databases. Furthermore, the author reports that current research by the Getty Research Institute continues to advance knowledge and drive new initiatives, such as the Online Scholarly Catalogue Initiative [7](OSCI), working in collaboration with museums and aimed at providing museum catalogues and more detailed information about artworks online and support for technological tools, for example, interactive technologies for examination of artworks. Gaehtgens (Ibid.) contends that the above features together with cross-referencing to other information and the abundance of materials available, organized in a manner to support art historical questions, whatever these might be, will inevitably lead to more collaboration between art scholars and may support better interpretation of materials and understanding of different cultural traditions.

Supporting art historical research is also the goal of the Ithaka S+R Research Support Services [8], who provide research support services for scholars, currently in the fields of, art history, chemistry and history. Their recent year-long case study in the field of art history, commencing in early 2013, highlights the changing practices of art historians and asks how services can adapt to provide effective support (Long & Schonfeld, 2014). In this study, seventy six interviews were undertaken with scholars ("who conduct some form of critical analysis of art in the course of their research"), through a series of visits at their respective institutions, including: the Getty Research Institute; Princeton University; the Seattle Art Museum; Swarthmore College; the University of Kansas and a short visit to the Yale University Haas Family Art Library. As might be expected, discovery and viewing of materials are being transformed by online tools, largely supplanting traditional finding and viewing aids of books, slides and photographic archive collections. However, findings from the study also reveal that limited online information about artworks; lack of high quality online digital images and disconnected repositories, hinder research. The latter identifies a significant problem for scholars, whose research spans collections from multiple institutions. Moreover, some participants noted a difficulty in knowing where to search and having to search many different locations. Additionally, many resources have no online presence, for example, collection of serials and archives and consequently valuable information is lost because these collections may be ignored. The authors propose that comprehensive metadata and semantic web technologies may address many of the identified problems. Setting aside the constraints and

[7] http://getty.edu/foundation/initiatives/current/osci/index.html
[8] http://www.sr.ithaka.org/research-publications/research-support-services-scholars

recommendations documented in the study unrelated to technology, further constraints for research include art researcher's lack of specialized technology skills: in organizing personal collections; accessing and manipulating images and note that training in the use of technology tools that can transform research allowing 3D modeling, Reflective Transformation Imagining and so forth.

Beaudoin (2014) identifies a gap in information systems design and development for professional image users. The author acknowledges the value of research into search and retrieval, but contends that understanding how and why images are used, will influence system design regarding the fundamental issues of the what, where and how information is sought and the specific characteristics of information that may be required. In this small case study of 4 archaeologists, 6 architects, 6 art historians and 6 artists, all twenty participants completed a survey and a semi-structured interview. The study documents that work processes of all participants involved downloading and collecting images for reference purposes and that all except the artist group used them in lectures and/or presentations. The artist group manipulated and printed the images and used them for inspiration and models for future artworks, whereas the art historians scanned the images and used them in PowerPoint presentations, but, surprisingly, only one art historian volunteered that he or she used images in research and publications. Findings from this study reveal that "to develop knowledge" was the most common and frequent reason across all four groups. This knowledge is acquired by examining an image, information exchange via the images with increasing knowledge through investigation of historical events; locations; cultural contexts; styles; people; themes and designs; and among art history students to develop visual and verbal-textual vocabulary. Other findings included: that images are frequently used as models for developing creative works, for communication, inspiration and cognitive recall and to a lesser degree for critical thinking development, proof, translating verbal information; engaging students; connecting with people; creating emotion; marketing and developing trust."

These studies provide valuable resources for this research and inform our research of how domain experts conceptualize their role and how they build knowledge of artworks (Chapter 4). The evidence from these (above) studies shows that while the work of art and its intrinsic qualities remain central to their research, domain experts consider external information essential to enhancing their understanding of the authentic nature of the work. Further evidence of this can be found in recent research that demonstrates an increasing interest in archival materials, particularly artist archives, as valuable sources for art historical research and discussion of these primary sources, continues below.

2.2.4. A note on Archives

Archives contain primary documents and artefacts created and accumulated by organisations and individuals during their lifetime, that are subsequently preserved as valuable sources of evidence of past activities by these entities (NAT UK, 2011). The earliest archives, from ancient

Greece and the Roman Empire were all but destroyed during the Great Invasions of the fifth, sixth and seventh centuries AD, and it was not until the thirteenth and fourteenth century that local and national administrations began to emerge, and with them the beginnings of archival repositories (Duchein, 1992). At that time, typical contents of these archival repositories comprised documents such as titles of land and property, commercial contracts, marriage deeds, last wills and so forth, and were preserved within municipal buildings, giving rise to the 'legal' status of the repositories. The author recounts that over the centuries the notion of archives as historical repositories only, evolved into one of repositories that also receive records for functioning institutions on a regular basis. Furthermore, debate about the arrangement of archives identified different approaches; either chronologically, by place names or according to the legal nature of the documents.

The principle of provenance or '*respect des fonds'* originally articulated in 1841 by Natalis de Wailly (archivist and historian) asserts that "all documents which come from a body, an establishment, a family or an individual from a *fonds* and must be kept together". Subsequently, the principle of '*respect for original order"* was identified, whereby records should be maintained in the same order in which they were received. These principles underpin archival theory developed in the 1850's, which informs the modern archival science, of today (Ibid.) and are incorporated into metadata standards developed for the domain, including ISAD(G) and EAD (discussed in 2.3)

Of particular relevance to this study, however are archives relating to the artist. Artist archives comprise unique, private, contemporaneous materials generated over the course of the artist's lifetime, and represent potentially fertile resources or raw materials for the art historian and curator (Breakell, 2015; Pyle, 2001; M. Reed, 2017; Sjöholm, 2014). According to M. Reed (2017) artist archives, combined with other public art historical resources "are foundational to creating the canon of the most valued or revered works of art and artists". Moreover, the authors also consider artist archives and impending archives of contemporary artists' studio materials as important stores of knowledge and sources of inspiration for the artist himself/herself. These authors also note their importance as sources of documentary evidence of their works. In agreement with the view of the potential weight of archival materials for the researcher, Schwartz and Cook (2002) assert that the archival document extends beyond that of "the bearer of historical content " in that it reflects "the needs and desires of the creator(s)" and the reason they were created. Koot and Koppenol (2016) note an increased interest in ephemeral materials in art historical research and that these represent rich sources of information, especially about relationships between artists, dealers and museums.

Ordinarily, archival materials reside in museum holdings and archivists facilitate access to the materials by domain specialists and those researching a work of art or the archives. Crookham (2015) however, shows that domain experts not only consult these materials when researching works of art, but in some cases the curator shares this knowledge to further enhance the visitor's experience, by integrating archival materials in fine art exhibitions. Furthermore, the author's exploration of this phenomenon revealed that curators distinguish between archives as historical and documentary evidence and archives as artworks in their own right. By way of example, the central theme of the Turner exhibition at the National Gallery in London in 2012, focused on the similarities and related themes in both Turner's and Claude Lorrain's artworks and demonstrated how Lorrain's art inspired and shaped some of Turner's art. While the exhibition demonstrated the importance of the Turner bequest to the gallery, it also revealed the relationship between both artists by including a room dedicated to archival materials, art objects and documentary evidence of that relationship. Interestingly, as the author notes, how this reflected the curator's interpretation of the art objects in this room, in that they were presented as documentary evidence as opposed to being considered as aesthetic objects. On the other hand, in the exhibition "Seduced by Art, Photography Past and Present" at the National Gallery in London Oct 2012 to January 2013, the curator placed photographs next to Old Master paintings, and the photographs were also viewed as aesthetic objects.

2.2.5. Discussion

The literature emphasised the complex nature of aesthetic objects, in relation to the different kinds of works of art and concerning the existence, identity and persistence conditions of these objects. It also confirmed the need for art experts or 'grounders' with their specialist knowledge of art and their expert practices, to evaluate and to preserve the intellectual value of works of art. At one end of the spectrum, the aesthetic empiricist or formalist art theory holds that the intrinsic qualities of visual art and art specialist knowledge of 'significant form' are all that is required to appreciate and evaluate a work of art. At the other extreme, the contextualist theory holds that facts outside of or beyond the intrinsic qualities of visual works are equally important. Art historical practices proposed by Panofsky and Gombrich, in addition to the identified need for provenance research, favours the contextualist approach. The studies presented show that art specialists consult a wide range of resources, including both manual or hard copy materials and computer systems to develop knowledge and improve understanding of these objects. This suggests an evolving research practice from the principle of "comparison being the methodological basis of the discipline of art history" (Brilliant, 1988) to one that equally values contextual and related information about a work of art, by way of example, for provenance information, critical reviews and archival materials. The studies presented also show that although more resources, such as journals, museum websites and art specialist web portals are available today, difficulty in knowing where to look for materials, and the fact that materials are missing and /or the limited information available online, highlights a shortfall in technology serving the needs of the art researcher. The research also shows an increasing interest in artist archives as potentially rich resources for art research, but much of these primary sources are not yet available online.

Organising and describing these complex objects is the subject of the next section, which explores the standards and tools developed for the domain, in relation to this.

2.3. Metadata and Resource Annotation

Exploration of the different kinds of artworks and the practices of art specialists concerning these works carried out in the previous section, revealed the complex nature of aesthetic objects. The problem of representing these complex objects and their relationships in a system that reflects the ontological commitment of art specialist's, needs to be examined. Mechanisms to describe the objects in such a way that captures the semantics of the domain, namely metadata schemata, vocabularies and classification systems are reviewed.

2.3.1. Metadata

The term *metadata* was coined by Philip Bagley in 1968 (Bagley, 1968, p. 26) in the context of programming languages, where one data element was used to represent data about another element (OED) *Meta* has its roots in the ancient Greek word μετα- denoting 'with', 'after', 'between': compare to Mycenaean Greek *me-ta* 'together with', which is perhaps the original sense in Greek. *Data* [9] derives from classical Latin *data*, plural of *datum*. *Metadata* is commonly defined as 'data about data' or 'data that describes and gives information about other data' (OED).

Traditionally used for cataloguing and indexing in libraries, metadata was generally regarded as a *descriptive* tool for fields in files; database components and document repositories (Sen, 2004; Shepherd & Pringle, 2002). However, *metadata* usage continues to be refined as new technologies and types of resources emerge, concurrent with changes in practises of organisations and users across different domains. In addition to identifying new categories of metadata: *administrative; descriptive; structural; preservation* and *technical* metadata, addressing primary concerns of the domain (Baca, 2008a, p. 9; Hyvönen, 2012, p. 35), a variety of metadata standards have evolved to support and inform administration of the diverse repositories in libraries, archives and museums (LAM's). Adopting appropriate metadata standards contributes to data consistency; resource discovery and retrieval; shareability; interoperability and data administration (Baca, 2008a; Edwards, Mayernik, Batcheller, Bowker, & Borgman, 2011; Otto, 2014; Pal, 2010; Riley, 2017). Furthermore, the metadata itself is recognised as an important asset for the long term preservation and management of materials (Baca, 2008a, p. 8).

Acknowledging the existence of comparable resources in other disciplines, this review is limited to resources specific to cultural heritage (CH). Nine sample metadata schemata in the CH domain are reviewed, including two conceptual models, reflecting the underlying rich and heterogeneous nature of the domain. The section concludes with a brief outline of other vocabularies that may be used for CH annotation.

9 In pl. Computing. The quantities, characters, or symbols on which operations are performed by a computer and which may be stored or transmitted in the form of electrical signals and held on recording media. Also (in non-technical contexts): items of information represented in digital form. (OED)

Table 2.7 Metadata schemata for cultural heritage, categorized according to Hyvönen (2012)

Schema	Name and Resources description	Start Year	Elements/ Sub elements	Core Elements (where stated)
Web Schemata				
DC	*Dublin Core:* used for simple and generic description of electronic resources.	1995	15/55	
VRA	*Visual Resources Association:* describes digital Images and works of Art and Culture (in digital format)	1996	19/37	c. 5/15
EAD	*Encoded Archival Description* Standard for archival and manuscript repositories.	1998	146	
Cataloguing Schemata – content standards				
SPECTRUM	*Standard ProcEdures for CollecTions Recording Used in Museums:* used to catalogue and manage materials in libraries, archives and museums.	1994	21/c.710	
CDWA	*Categories for the Description of Works of Art:* describes works of art, architecture, other material culture, groups and collections of works, and related images	1996	31/c. 504	c.45
&	'Lite'/reduced set of CDWA categories	2005	22	22
ISAD(G)	General International Standards for Archival Description.	1994	26	6
Conceptual Schemata				
			Classes	**Properties**
CIDOC CRM	Conceptual Reference Model: High Level Ontology for CH materials in Libraries, Archives and Museums (LAM's)	2006	93	164
EDM	Europeana Data Model Designed for the Europeana project, describing semantics of materials in LAM's in Europe.	2011	16	78

2.3.1.1. Dublin Core

Dublin Core was introduced in Dublin, Ohio in 1995 as a simple standard for generic description of networked resources (Weibel & Lagoze, 1997) and is the result of collaboration involving experienced, international and multidisciplinary team comprising librarians, digital library researchers, networking professionals, text mark-up specialists and content specialists. The standard comprises fifteen core elements: *Creator; Title; Format; Identifier; Type; Contributor; Date; Subject; Relation; Coverage; Publisher; Language; Description; Source* and *Rights.* DC is widely adopted and was approved as the NISO standard (Z39.85) in 2001. However, a survey of 141 experienced cataloguing and metadata professionals identified problems and confusion about some of the elements (Park & Childress, 2009) with the potential inconsistency and inaccuracy in semantic interoperability across repositories.

Since the late 1990's "it has evolved in the context of a Dublin Core Metadata Initiative (DCMI) incorporated as a not for profit organization hosted at the National Library board of Singapore." Today it is a set of standards that includes an Abstract Model (with 22 classes), a Framework for Application Profiles, a Type Vocabulary and includes 55 terms of 4 types: properties; classes; datatypes; and vocabulary encoding schemes. *Classes* are groups of resources sharing common characteristics, for example Class Agent might contain artists or creators. *Properties* refer to the core attributes for resource description, for example, dc:title, dc:creator or dcterms:title). *Datatypes / Syntax Encoding Schemes* (SES) specify how a value has to be structured, for example: dcterms:date syntax may be determined by SES W3CDTF. Vocabulary Encoding Schemes identify controlled vocabularies constraining the values of a property, for example: dcterms:subject may be restricted to a concept defined in AAT (Art and Architectural Thesaurus) vocabulary.

DC operates on the 1:1 principle: whereby one record describes the digital surrogate, for example, image1.jpeg while another record describes the artwork from a cultural heritage perspective.

While DC appears in the Web Schemata Category in the above table, the scope of DC has been extended to include in principle, any object whether electronic, real-world or conceptual, in an abstract model.

2.3.1.2. VRA: Visual Resource Association

VRA was introduced in 1996 as a data standard for the cultural heritage community. Upgraded to Version 4 in 2007 by the VRA Standards committee of the Library of Congress, USA (VRA, 2007) it comprises 19 metadata elements to describe works of visual culture: *Work, Collection; or Image (Id);Agent; Culturalcontext; Date; Description; Inscription; Location; Material; Measurements; Relation; Rights; Source; StateEdition; StylePeriod; Subject; Technique; Textref; Title* and *Worktype.*

In total, VRA contains 37 elements (including sub elements) along with 9 global attributes that may be applied to any element, as appropriate. However, a minimum of 5 elements is required to define a record: *Worktype*; *Title*; *Agent*; *Location* and *Date*. The global attributes provide access to external data, for example, the *vocab* and *refid* elements may be used to store the name of an external vocabulary and the record identifier in the vocabulary which can then be used to access data from the specified vocabulary; the *href* element denotes a hyperlink to another electronic resource.

VRA may describe a *work*, *image* or a *collection*. A *work* is a unique entity an *image* is a digital representation of a work and a *collection* describes an aggregate of work or image records, conceptually or physically arranged for the purpose of cataloguing or retrieval, for example an archival entity. VRA also adheres to the 1:1 principle of Dublin Core through the Relation element. The Relation element is used to depict the relationship types, work-to-work; image-to-work; image-to-collection/ work-to-collection. The unique identifier of the main record must be specified in the individual work and image records in order to establish the link.

VRA development has been influenced by CCO standards (Cataloguing Cultural Objects) "specifically in the differentiation of data values for *display* and *indexing*".

2.3.1.3. EAD: Encoded Archival Description

EAD was introduced in 1998 as a standard for encoding archival and manuscript repositories in a networked (online) environment. The project was originally undertaken in 1993 by the University of California, Berkley, in response to the need for universal access to archival materials (Pitti, 1997). Developing online resources (in XML format) offered greater flexibility for information gathering and communication with researchers.

EAD Version 3 consists of 165 tags/metadata elements for describing the repositories. By way of example, <archdesc> Archival Description where archdesc is the machine readable tag and Tag and *Archival Description* is the element name. An additional 85 attributes may be used as qualifiers to the elements, where the attribute values differ, depending on the context / tag name being qualified. For example, the 'unitdate' is qualified as being 'inclusive' and 'mormal' or the standardised form of the content.

```
<[tag] [attribute1]="[value1]" [attribute2]="[value2]">
<unitdate unitdatetype="inclusive" normal="1937/1992">1937-1992
</unitdate>
```

For example, the NORMAL attribute may limit the element value to a controlled vocabulary: the COUNTRYENCODING attribute restricts a value to standards for country encoding iso8601.

Archival records are stored in "unfolding hierarchical levels". LEVEL, a general attribute qualifying the <archdesc> element denotes the context of the material in the collection, where a collection is described in descending order of *collection; fonds; class; recordgrp; series; subfonds; subgrp; subseries; file; item;* or *otherlevel*. The <archdesc> element contains a description appropriate to the level of the material being described. Higher level descriptions are presented in subordinate materials, before the description (<dsc>) of the lower level item.

14 of the 85 attributes may be used to establish *links* from items in the archive collection: the ID attribute identifies a material so that it can be referred to and the Parent attribute points to the container in which the current item being described is held. A link to remote materials may be established using the ENTITYREF and HREF attributes, for example,

```
<extptr linktype="simple" entityref="phyllis" title="Image of Person Name"
        actuate="onload" show="embed">
```

EAD is maintained by the Library of Congress since 1998 with the Society of American Archivists responsible for ongoing oversight of the standard. A third version EAD3 was announced in august 2015. Recent research (Francisco-Revilla, Trace, Li, & Buchanan, 2014) uncovered previously unidentified problems that the authors claim have significant impact on the ability to visualize this data. Their work analyzes a set of 8729 finding aids aggregated by the Texas Archival Repository Online (TARO) using VADA, a visual analytic tool for finding aids. The authors assert that many of the problems identified are based on a mismatch between the needs of those arranging collections and encoding finding aids, and those involved in aggregating and visualizing the sets of finding aids. Since this study was published in 2014, it is not clear if these problems are resolved in EAD 3.

2.3.1.4. CDWA: Categories for the Description of Artworks

CDWA describes the content of art databases and was designed to satisfy the information requirements of those who catalogue works of art, domain professionals and scholars researching art. Introduced in 1995 and maintained by the Getty Research Institute (GRI), it describes and enables access to information about works of art, architecture, other material culture, groups and collections of works and related images. Defined as a metadata elements set or schema (Baca, 2008a, p. 10; Baca & Harpring, 2014 -Revised online 2016) it contains 31 categories, comprising c. 540 metadata elements. However, only a small subset of categories (14) with c.35 elements are considered *core* to uniquely identify and describe a work.

CDWA also emphasises standards for data values through the use of controlled vocabularies and thesauri resources (ibid. p.58, 76) and for descriptive practises: promoting consistency and accessibility to works of art across multiple databases. Description of artworks in CDWA is informed by guidelines for Cataloguing Cultural Objects (CCO) (Baca et al., 2006; Coburn, Lanzi, O'Keefe, Stein, & Whiteside, 2010)

2.3.1.5. CDWA Lite

CDWA Lite is an XML schema to describe core records for works of art and material culture based on CDWA and CCO (Coburn et al., 2010). Comprising twenty two elements, nineteen elements describe the artwork and three are for administrative metadata. (Object/Work Type; title; Display Creator; Indexing Creator; Display Measurements; Display Materials/Techniques; Indexing Materials/ Technique; display State/Edition; Stype; Culture; Display Creation Date; Indexing Dates; Location/Repository; Indexing Subject; Classification; Description/Descriptive Note; Inscriptions; Related Works; Rights for Work; Record and Resources. It is regarded as a simple and efficient model for machine-readable records, enabling even the smallest museums to contribute to union catalogues (ibid).

2.3.1.6. SPECTRUM

Introduced in 1994, SPECTRUM (*Standard ProcEdures for CollecTions Recording Used in Museums)* is an international standard for collections management in museums. The initial release was the product of over 20 years work by the Museum Documentation Association, and three years in cooperation with over 100 practicing museum professionals (Shepherd & Pringle, 2002). SPECTRUM comprises c. 700 metadata elements applied to its 21 documented procedures, from the initial receipt or acquisition of an object/collection through all procedures up to and including disposal/de-accession of an object, with documented guidelines available for all procedures. However, museums will not necessarily implement all of the 21 procedures.

In response to changes in professional practices, the standards were revised as Version 5, released in September 2017. Compliance with the standards accreditation scheme requires museums to implement the nine core procedures identified in Version 5 [10]. Furthermore, these procedures may be applied to manual or paper based systems, to computerised systems or a combination of both.

2.3.1.7. ISAD(G)

General International Standard Archival Description ISAD(G) was introduced in 1994 as a set of standards for the preparation and maintenance of archival descriptions (ICA, 2000). The standard identifies and defines seven categories of information comprising 26 elements to describe an archival entity, six of which are mandatory for all records described. The mandatory elements include *Reference code; Title; Creator; Date(s); Extent of the unit of description; and Level of description*. The standards may be adapted as appropriate by each nation (Appendix I).

ISAD(G) was developed by the International Council on Archives (ICA) Ad Hoc Commission on Descriptive Standards (ICA/DDS), which became a permanent committee at the ICA International Congress on Archives in Beijing, China, in 1996. The Committee on Descriptive Standards (ICA/CDS) undertook the revision of ISAD(G) (Ottawa, 1994) as its primary task in its program for 1996-2000. ISAD(G) archival records are stored in hierarchical levels: *Fonds; Sub-fonds; Series; Sub-series; File* and *Item*, Fonds representing the whole/aggregate of the records and Item representing an individual component at the bottom level of the archival entity. Collection descriptions are provided for each of the elements. To avoid repetition of descriptions, the standard recommends that descriptions provided should be appropriate to the level in the hierarchy and that all higher level descriptions are available to all lower level materials.

Allied to ISADG, the International Standard Archival Authority Record for Corporate Bodies, Persons and Families ISAAR (CPF) (ICA, 2004), is required to record authorities and then to link these authorities to ISAD(G). Authority records are used to describe, a corporate body, person, or family as units within an archival descriptive system; to control the creation and use of access points in archival descriptions and to document relationships between different records creators and between those entities and the records created by them and/or other resources about or by them (Ibid.)

Unlike EAD, ISAD(G) does not prescribe formats for presentation of the contents.

[10] http://collectionstrust.org.uk/spectrum/spectrum-5/ Last accessed 06 Mar 2018

2.3.1.8. CIDOC CRM: Conceptual Reference Model

(Comité Internationale pour la Documentation (The International Committee for Museum Documentation, Conceptual Reference Model)

CIDOC CRM (CRM) is a high-level ontology for the cultural heritage (CH) domain (Doerr, 2003): the latest version became available in December 2014, as ISO21127:2014. [11]

Developed by the International Committee for Museum Documentation Standards Group (CIDOC) of the International Council of Museums (ICOM), it was first published in 2006. Development of CRM however dates back to 1996, when the decision was taken to move from the Relational Data Model to an Object Oriented or Conceptual Reference Model (Crofts & Reed, 1996). Thereby also extending the scope from that of museum objects and their provenance to providing an "intellectual and conceptual framework formally describing and integrating complementary domain-specific models". (ibid)

Described as "an ontology for cultural heritage, it reflects the 'world view', of the underlying database applications and documentation systems that are used by cultural heritage institutions.... , whereby CRM aims to clarify the logic of what cultural heritage institutions do in fact document." (LeBoeuf, Doerr, Ore, & Stead, 2013). CRM facilitates the description of explicit and implicit concepts and relations relevant to the documentation of cultural heritage, in a formal language"[12]. It serves as a guide for the transformation of disparate metadata models used by memory organisations into a form compatible with a foundational semantic ontology model [13] (Hyvönen, 2012, p. 44). Yet, CRM does not specify the individual elements that must be used to describe CH materials and surrogate digital data.

The intended scope of CRM is defined as "the exchange and integration of heterogeneous scientific documentation related to museum collections" (Ibid.) CRM comprises 93 classes and 164 properties for organising and describing cultural heritage and relationships therein. It is classed as an event centric model (Doerr, 2003; Hyvönen, 2012, p. 44). For example, *events* (such as acquisition; modification; birth; death and so forth) are mapped to an appropriate event sub-class and associating the event with one or more concepts/persons in a sub class of Persistent Entity and to a Place and Time-Span. Furthermore, multiple instances of the same event may be specified in order to preserve contextual information; for example, a place name may change over time; the correct attribution of a work of art may be discovered at a later point in time, creation date or birth dates may be revised as new information becomes available: all of these events are preserved in the model.

2.3.1.9. EDM: Europeana Data Model

[11] http://www.iso.org/iso/catalogue_detail?csnumber=57832 ISO 21127:2014 establishes guidelines for the exchange of information between cultural heritage institutions: museums, libraries, and archives. Documentation is now available.. Last accessed 23 Nov 2015
[12] http://www.cidoc-crm.org/index.html Last accessed 23 Nov 2015
[13] http://www.cidoc-crm.org/comprehensive_intro.html Last accessed 23 Nov 2015

EDM was introduced in 2011 as a top-level ontology for the Europeana portal [14] (Section 2.4). EDM is “aimed at being an integration medium for collecting, connecting and enriching the descriptions provided by Europeana content providers” [15] such as libraries, archives and museums (LAM’s).

Inherent problems of Europeana Semantic Elements (ESE, the predecessor of EDM was launched in 2008) [18] (Doerr et al., 2010; Peroni, Tomasi, & Vitali, 2013), compelled redesign of the model. Consequently, work began in 2009 on a semantic web based framework. Similar to CIDOC CRM, EDM, is a formal specification of domain concepts on which diverse schemata representing the rich and heterogeneous cross-domain collection metadata in LAM’s are transformed (Hyvönen, 2012, p. 43); (The transformation or mapping is currently performed by content providers EDM Primer p.18). EDM designers have stated that they have addressed some of the shortfalls of ESE, namely, EDM eliminates prescription of a standard metadata set across all contributors; it preserves the relationships between data received from content providers; multiple records for an object may be provided; it now distinguishes between the cultural material record and the digital record (for example a .jpeg of a painting).

Constructed as a top-level ontology EDM comprises 16 classes and 78 properties. Classes being “groups of things with common characteristics” and Properties as “equivalent to a data element”, [16] Some classes and properties are EDM-specific whereas others reuse existing namespaces such as RDFS, SKOS, ORE and DCAT (Chapter 3). For example, the edm:Agent class consists of “persons or groups who have the potential to perform intentional actions for which they can be held responsible” such as artists, authors, museum and so forth and the property edm:happenedAt “associates an event with the place where the event occurred”. EDM also addresses interoperability between EDM and other models, for example the edm:Agent class is equivalent to E39_Actor in CIDOC CRM and edm:happenedAt property is equivalent to P7_took_place_at in CIDOC CRM.

Although EDM was specifically developed for Europeana, it has been adopted and extended by the Smithsonian American Art Museum project to publish its data on the semantic web (Szekely et al., 2013) being perceived as a rich and easily extensible domain ontology. A recent study (Zapounidou, Sfakakis, & Papatheodorou, 2014) reported that mapping BIBFRAME [17] to EDM is “achievable without significant loss of semantics”, however, it also recommends future study to resolve some mapping issues and because BIBFRAME is an emerging technology.

2.3.2. Discussion on Metadata Standards

The standards reviewed are the result of more than twenty years research and collaboration among scholars across multiple disciplines of cultural heritage (CH), visual resource curators

[14] Portal : http://europeana.eu Last accessed 23 Nov 2015
[15] Primer: http://pro.europeana.eu/documents/900548/770bdb58-c60e-4beb-a687-874639312ba5 Visit 23 Nov 2015
[16] http://www.w3.org/2001/sw/wiki/Semantic_Web_terminology Last accessed 23 Nov 2015
[17] http://www.loc.gov/bibframe/ Last accessed 23 Nov 2015

and specialists in information systems. As such, they reflect the specialist practices and body of knowledge, across the CH community. Evolving from the traditional *descriptive* role, the significant impact of metadata standards on the domain is obvious: contribution to data consistency, improvement of workflows, collaboration, administration, discovery, dissemination and integration of CH materials. On the other hand, the increasing specialization and sophistication of standards is motivated by the expressed needs of CH specialists.

The variety of existing standards results from evolvement over long periods of time; they are created and revised as new technologies emerge; as new resource types emerge; the standards reflect different practices of CH communities worldwide; adoption and adaptation in recognition of the benefits and opportunities presented. Furthermore, the review reveals the heterogeneous nature of CH materials and consequently, selection of appropriate standards for cross-domain materials is essential. For example CDWA Lite facilitates dissemination of artworks to union catalogues, yet it has been expanded by the German Museums Association, resulting in Museumdat, (an XML schema incorporating natural and cultural history collections) (Coburn et al., 2010); the role of CIDOC CRM is mediation and integration of CH cross-domain materials on the semantic web, whereas ISAD(G) describes materials in an archival entity which in turn may be encoded in EAD for dissemination online, and CDWA describes works of art and may not adequately describe the hierarchical and descriptive complexity of archival entities.

Interestingly, the literature reveals the converging of standards as technologies and their adoption by CH communities matures. Beyond this review, the international standard ISAD(G) (2.2.3.7) has largely influenced the 2013 version of DACS (Describing Archives: a Content Standard) in the USA [18] which in turn replaces a previous US standard Archives, Personal Papers, and Manuscripts (APPM). One observes also that more than one standard exists for the same category of materials: SPECTRUM and CDWA (with CCO content guidelines), both developed in tandem over decades in the UK and USA, respectively. Adopted by their respective communities, they continue to evolve in support of the communities. Similarly, continual development of CDWA, SPECTRUM and EAD supports the large community adoption, of those standards.

The limited number (26) of elements in ISAD(G) compared to the granularity of EAD may raise problems whereby mapping elements from ISAD(G) to EAD necessitates splitting some data items to accurately complete the mapping and many EAD elements are not filled. Moreover, ISAD(G) lacks elements to establish links with external entities (beyond textual description), for example, to link an archival collection or any of its element to an object or an agent, for example Jack B Yeats. A separate standard, the International Standard Archival Authority

[18] http://www2.archivists.org/groups/technical-subcommittee-on-describing-archives-a-content-standard-dacs/dacs Last accessed 23 Nov 2015

Record for Corporate Bodies, Persons and Families ISAAR (CPF) is required to record authorities and then to link these authorities to ISAD(G). EAD comprises elements and qualifiers to establish such remote links.

The above standards are incorporated into the major software systems for museum collections [19] and Archives [20] excluding by definition, CIDOC CRM and EDM. As noted in the introduction, a limited set of metadata is reviewed: omitted standards include: FBBR (Functional Requirements for Bibliographic Records) and the Object oriented version; MARC and MARC 21 (MAchine-Readable Cataloging), the Library of Congress standards which have served the bibliographic for over forty years and BIBFRAME (Bibliographic framework) its' impending replacement.

Regarding Dublin Core (DC) a survey of 141 experienced cataloguing and metadata professionals identified problems and confusion about some of the elements (Park & Childress, 2009) with the potential inconsistency and inaccuracy in semantic interoperability across repositories.

These metadata standards inform the design of our computational representation that incorporates the knowledge of the curator. The researcher selected CDWA, VRA in conjunction with CCO (Cataloging Cultural Objects) that describes the content standards of the database, to design the relational database schema for the works of art (described in Section 3.5). The source archival documents harvested in this study are described according to the ISAD(G) standard and the researcher selected ISAD(G) and EAD standards to design the relational database schema for the archival materials (described in Section 3.6).

2.3.3. Resource Annotation

Resources for annotation of cultural heritage materials and classification systems play an important role in a computational representation of aesthetic objects. These knowledge bases provide standard terms that may be used to catalog, describe and classify works of art and can assist in the search and retrieval of cultural heritage materials (Harpring, 2010b). Furthermore, these resources may enrich the representation of artworks by providing additional information beyond that which may be stored in an institution's local system (Ibid.).

Acknowledging the existence of comparable resources in other disciplines, this review is limited to resources specific to cultural heritage (CH), comprising the Getty vocabularies, thesauri of content values for metadata items for cultural heritage materials and two classifications system, Iconclass, a classification system for the description and analysis of subject matter of visual images and the Index of Christian Art at Princeton University, an iconographic register of the subject matter of medieval art.

2.3.3.1. The Getty Vocabularies

[19] http://www.collectionstrust.org.uk//collections-link/collections-management/spectrum/choose-a-cms Visited 23 Nov 2015

[20] http://www.clir.org/pubs/reports/spiro/spiro_Jan13.pdf Pg. 42-43 Visited 23 Nov 2015

The Getty Research Institute is a program of the J. Paul Getty Trust "dedicated to furthering knowledge and advancing understanding of the visual arts and their various histories" [21] Therein, the Getty Vocabulary Program (GVP) maintains structured vocabularies for cataloguing and retrieval of information about works of art and as general research tools (Harpring, 2010a)

Vocabulary development commenced in the late 1970's in response to the need for consistent cataloguing and efficient retrieval (ibid). The existing Library of Congress Subject Headings did not meet the specialist needs of the arts community (van Hooland et al., 2013; Weinberg, 1995), providing the impetus for the Getty to begin work. The GVP produced three vocabularies:

I. Art and Architecture Thesaurus (Wibisono et al.)

II. Union List of Artist Names (ULAN)

III. Thesaurus of Geographic Names (TGN)

Table.2.8 Development of the Getty Vocabularies (Isemann & Ahmad, 2014), updated to reflect current volumes.

	Work		Size, May 2011		Size, June 2014		Size, July 2017		Growth (2011-2017)	
Vocabulary	Began	Published	Records	Names	Records	Names	Records	Names	Records	Terms
AAT	Late 70's	1990	34,880	245,530	51,470	268,650	60,000	375,000	72%	53%
ULAN	1984	1994	202,720	638,900	248,820	638,818	300,000	720,600	48%	13%
TGN	1987	1997	992,310	1,711,110	1,431,380	2,035,195	2,550,000	4,100,000	157%	140%

A thesaurus is a controlled vocabulary of terms organized systematically to represent equivalent, homographic, hierarchical and/or associative relationships between subjects and may also identify synonyms, antonyms and homographs for the terms (Calaresu & Shiri, 2015). The Getty vocabularies are thesauri that "provide terminology, relationships and other information about the concepts, artists, places and objects important to disciplines specialising in art, architecture and material culture" (Harpring, 2010b).

Organised in hierarchical structures of primary divisions or *facets,* these in turn contain *hierarchies*. Information is stored in records, the focus of a record being a concept (depending on the vocabulary), identified by a unique *Subject ID.* Position within the hierarchy is maintained by a *Parent ID*. *Relationship types* defined within the vocabularies include:

- *Hierarchical* relationships: either Whole/part for example, members of a corporate body; *Genius/species*, for example, bronze is a *type of* metal; (Harpring, 2010b, p. 38) or

[21] http://www.getty.edu/research/institute/ (Last accessed 23/11/2015)

- *Instance* relationships: when used with proper names, for example, Jack B. Yeats is an *instance* of the 'Persons, Artists' facet in ULAN.
- *Associative* relationships: exist between concepts that are conceptually close. For example, 'Final Neolithic Age' and 'Early Bronze Age' (Ibid. p 42); or between an artist and other notable persons. *Associative* relationships are named and reciprocal. For example, if A is associated with artist B by the named relationship 'Student of', the named relationship 'Teacher of' is recorded for artist B (Ibid. p.47).
- *Polyhierarchical* relationships: linking one record to multiple parents also exist. For example, Siena (Sienna), a city in Tuscany in the modern world, is also part of the ancient confederation of Etruria. (ibid. p. 41): the non-preferred parent qualifier is *N*.
- *Equivalence Relationships* , Multiple *terms* and other concepts may be linked to a concept (Fink, 1999): terms with similar meanings / synonyms or lexical variants are considered Equivalence Relationships.

A qualifier field distinguishes between homographs. Terms may be expressed in the singular, plural, various forms of speech, variant spellings and multiple languages. A *preferred* (descriptor) *term* is required: the default being the American English term.

Currently available in database and XML format, the 2014 Getty project aims to publish the vocabularies as Linked Open Data. The vocabularies are compliant with ISO & NISO standards for thesaurus and controlled vocabulary construction. (Harpring, 2010a).

2.3.3.1.1. AAT: Art and Architecture Thesaurus

AAT is a controlled vocabulary of *generic* terms and information about concepts used to describe art, architecture, decorative arts, material culture and archival materials (Harpring, 2010b, p. 52)

AAT hierarchical structure comprises seven primary divisions or *facets.* These include; *Associated Concepts, Physical Attributes, Styles and Periods, Agents, Activities, Materials, Objects* and *Brand Names*. [22] Facets contain members which in turn may be further subdivided. For example, the *Physical Attributes* facet contains four hierarchies for *Attributes and Properties, Conditions and Effects, Design Elements* and *Colour,* whereas the *Materials* facet contains only one, *Materials.*

Position within the hierarchy of concepts is maintained through the use of a *Parent ID*: the parent representing the broader context, examples being, *Aegean architectural styles* is the broader context of *Doric (architectural style)*; *Late Baroque, High Baroque* and *Earlier Baroque* are narrower concepts of *Baroque*, for which *European* is the broader context, itself being a member of *Styles and Periods* hierarchy, within in the *Styles and Period facet*. An example of a

[22] http://www.getty.edu/research/tools/vocabularies/aat/index.html

Polyhierarchical relationship may be seen in the concept jade: *jade* has two broader contexts, *metamorphic rock* and *gemstone.* (Harpring, 2010b)

Data is stored in records: the primary fields include Subject ID and Record Type (concept; guide term; hierarchy name; facet; label; note and language for the note). The majority of records are of type concept. Multiple different terms may be associated with a Subject ID, considered as *equivalents* or *synonyms.* Each term is stored in a record identified by a unique Term ID. Data defined for a Term record include: type (one term must be the preferred term or descriptor) language; language status, dates of usage of the term; whether the term is historical or current; vernacular flag; part of speech flag (whether plural/singular, noun and so forth); preferred flag for a given language and a qualifier may be used to distinguish homographs. An example AAT record for the concept 'Expressionist (style)' is included in 7.1.10.2 (at Appendix H: Getty Vocabularies).

Work on AAT began in the late 1970's and was informed by the Library of Congress Subject Headings and Medical Subject Headings (Weinberg, 1995). It was published in 1990 by Oxford University Press. The complete second edition was published in 1994 in electronic form and in print by Oxford University Press. [23]. From 1997 it is published in automated formats only, in both a searchable online Web interface, downloadable data files available for licensing and as Linked Open data from 2014[1]. In 2006 the AAT was listed as one of the best free reference websites (Pappas et al., 2006)

2.3.3.1.2. ULAN: Union List of Artist Names

ULAN is a controlled vocabulary of *proper nouns*: comprising individual artists and those associated with art: corporations of artists and architects and non-artists, together with information about them. "Temporal coverage of the ULAN is from Antiquity to the present and the scope is global" (Harpring, 2010b)p.62).

Even though the structure is relatively flat ULAN it is constructed as a hierarchical database (ibid. p.62). It comprises five primary divisions or facets. These include; Persons, Artists; Corporate Bodies; Non Artists; Unidentified Named People and Unknown People by Culture. [24] Facets contain members (grouped in alphabetical order) which in turn contain records.

Position within the hierarchy is maintained through the use of a *Parent ID*: the parent representing the level above. Examples being, a studio or workshop in the Corporate Bodies facet may have members depicting *part-of relationship*s; Pablo Picasso is an *instance* of the <P> group within 'Persons, Artists' facet. An example of a *Polyhierarchical relationship* from the ULAN web page is the Foundation of Documents of Architecture (corporate body) having two broader contexts: 'Architectural Drawings Advisory Group' and 'The National Gallery of Art'.

[23] http://www.oxfordartonline.com.elib.tcd.ie/subscriber/article/grove/art/T000037?q=AAT&search=quick&source=oao_gao&pos=1&_start=1#firsthit

[24] http://www.getty.edu/research/tools/vocabularies/ulan/index.html

Associative relationships include for example, apprentice of; colleague of; patron; benefactor; student of: where relationship names are defined in a controlled list.

Data is stored in records: the primary fields include Subject ID and Record Type (this can be facet name, guide term, label; note and language for the note). The majority of records focus on artists/creators. Multiple different terms may be associated with an artist: considered as *equivalents* or *synonyms.* Each term is stored in a record identified by a unique Term ID. Terms describing the artists include: artist name; nationality; roles; gender; birth and death dates; events; associative relationships; biographies and descriptive notes. A preferred term must be specified for artist name, nationality and role. Data defined for a Term record include: type (one term must be the preferred term or descriptor) language; language status, dates of usage of the term; whether the term is historical or current; vernacular flag; part of speech flag/whether plural/singular, noun and so forth; preferred flag for a given language.

An example ULAN record for the artist Jack B Yeats is included in (Appendix H 7.1.10.1)

ULAN is the second vocabulary created by GRI. Work commenced in 1984, by merging authority files of seven Getty projects and two Getty programs (Fink, 1999; Harpring, 2010a). The first edition of the ULAN was published in hardcopy in 1994 [25] and machine-readable files. From 1997 it is published in automated formats only, in both a searchable online Web interface, downloadable data files available for licensing and is "planned to be available as Linked Open data from 2015".

2.3.3.1.3. TGN: Thesaurus of Geographic Names

TGN[26] is a controlled vocabulary of *place names* together with information about each place. The types of places included are natural physical features (continents; mountains; oceans; rivers; forests; submerged islands and former continents) and political/administrative entities (inhabited places and nations) (Harpring, 2010b)p.59. "Temporal coverage is from prehistory to the present and the scope is global" (ibid).

TGN hierarchical structure comprises two divisions or *facets: World* and *Extraterrestrial Places*. The *World* contains *Continents*, which contain Physical Features and Political entities. Political Entities for example may contain, Regions, Historical regions and Nations. The above examples represent broader context *whole/part relationships*. *Polyhierarchical relationship* may exist also: if a place has different physical and political parents; if it occupies two different countries or jurisdiction is disputed by different nations, then the place will appear under each of higher level divisions. In contrast to the above, where a physical feature (for example a river or mountain) or an historical region crosses more than one national boundaries then it will be placed in the next highest level division only. For example, the Alps appear under the broader context Europe only. Likewise the Roman Empire is placed directly under the World facet.

[25] (*Union List of Artist Names*. Project manager, James M. Bower; senior editor, Murtha Baca. New York: G.K. Hall, 1994)

[26] http://www.getty.edu/research/tools/vocabularies/tgn/about.html

Associative relationships may denote important historic ties between two places (ibid. p.45). Relationships are named (for example, formerly related to; ally of; capital of) and must be selected from a controlled list. Furthermore, the relationships are reciprocal.

Data is stored in a record: the primary fields include Subject ID and Record Type (Administrative; Physical; Both, hierarchy name; guide term and facet). The focus of each record is a place, identified by a unique numeric *Subject ID*. Multiple terms may be associated with a Subject ID: considered as *equivalents* or *synonyms.* Each term is stored in a record identified by a unique Term ID. Terms may be expressed in the vernacular, English names, variant names in other languages and historical names. A *preferred* name must be specified. Other flags used indicate various characteristics of the term: for example if it is an Official name, a Pseudonym, a Site name and so forth; a historical flag denotes if the term is current, historical or both. The language of the name is selected from a controlled list. Start and End dates may be specified for a term denoting the time span in which the term was used. Dates are in the proleptic Gregorian calendar with BCE dates expressed as negative numbers. If the relationship between place and date extends to the current time the End date is set to 9999.

TGN is the third vocabulary created by the GRI. Work on the TGN began in 1987. "Its development was informed by an international study completed by the Thesaurus Artis Universalis (TAU), a working group of the Comité International d'Histoire de l'Art (CIHA), and by the consensus reached at a colloquium held in 1991, attended by the spectrum of potential users of geographic vocabulary in cataloging and scholarship of art and architectural history and archaeology." (Harpring, 2010a). The TGN was first published in 1997 in machine readable format. From 1997 it is published in automated formats only, in both a searchable online Web interface and downloadable data files available for licensing, and is available as Linked Open Data from 2014. TGN continues to evolve as administrative areas change and archaeological, world heritage and other historical sites are added.

2.3.3.1.4. CONA: Cultural Objects Name Authority ®

CONA is a controlled vocabulary of names of works of art and architecture together with information about each one. The scope of CONA extends to "cultural works such as architecture and moveable works such as paintings, sculpture, prints, manuscripts, photographs, performance art, archaeological artifacts" (Harpring, 2010b) p.65, cataloged in scholarly literature, museum collections, visual resources collections, archives, libraries, and indexing projects with a primary emphasis on art, architecture, or archaeology." [27] "Temporal coverage is from prehistory to the present and the scope is global" (ibid), therefore, it may contain works that no longer exist.

CONA is the fourth vocabulary created by the GRI. Officially announced in 2008 (Baca, 2008b) although discussions began in 2005, coinciding with the development of CCO (*Cataloging Cultural Objects*) and CDWA (*Categories for the Description of Works of Art)* (Harpring,

[27] http://www.getty.edu/research/tools/vocabularies/cona/about.html

2010a). The aforementioned inform the data items required to describe a work of art and are outlined in the metadata section of this document. By 2010 early stage development had commenced: the initial dataset comprising moveable works from other Getty projects (ibid.)

A review of the CONA web page in November 2014 reveals that only a total of 1,040 records exist: these comprise 65 built works and 975 moveable works. The web page states *"It is hoped that the contributions to CONA from the user community will commence in 2011* and it is planned to be available as Linked Open data from 2015"

2.3.3.2. Discussion on the Getty Vocabularies

The product of research and collaboration among scholars across multiple disciplines such as art and architectural historians, architects, librarians, visual resource curators, archivists, museum personnel, and specialists in thesaurus construction, [28] the vocabularies preserve a considerable body of *specialist*, *authoritative* information for the cultural heritage domain, worldwide. Constructed as thesauri, the rich repositories of data and relationships types serve the needs of a diverse audience: researchers and cataloguers in cultural heritage and end user search and discovery.

Complementing the above strengths, the formal structure of the vocabularies allows for continual expansion and revision: the addition of new subjects or terms and the establishment of a variety of relationships between data, meeting possible variations required by different cultures. Table 5 shows the continued expansion as specialists worldwide adopt and contribute to the vocabularies; with recent contributions to AAT from Latin America (Nagel & Miller, 2013) and Taiwan (Baca, 2013)

The vocabularies are specialized, multilingual, with a global content and reach and with temporal cover from antiquity (ULAN) and prehistory (TGN, CONA) to the present. Recognised as the domain standard, they have been integrated to software systems in the domain, for example, AAT and TGN are integrated to The Museum System (TMS) [29] and AAT is supported in Adlib [30] and EmbARK, promoting data consistency, sharing and discovery across the cultural heritage domain.

A disadvantage may be that while the vocabularies are available online, a license is required to download and implement off-line, whereas others for example, LCSH (Library of Congress Subject Headings) and GeoNames are freely available. Moreover, the downloadable versions are only updated once a year, in contrast to GeoNames which is available daily. (Although appreciating the fact that GeoNames is crowd sourced and not an authoritative, controlled

[28] http://www.getty.edu/research/tools/vocabularies/contribute.html
[29] http://www.gallerysystems.com/tms
[30] http://www.adlibsoft.com/products/museum-collection-management-software

repository) Perhaps the Getty licensing situation may change given the Linked Open Data project that is underway.

Furthermore, users assume confidence in the accuracy of data in an authoritative repository. However, inspection the ULAN record for artist Jack B. Yeats shows confusion in birth and death locations between the artist and his father and the Note section is inaccurate. Also, for the poet William Butler Yeats the ULAN preferred term is ‘painter’.

Finally, the move to Linked Open Data is a positive step and contributes further to the adoption of the vocabularies, on the Semantic Web.

2.3.3.3. Classification systems

This section describes two classification systems developed iconographic analysis and description of images or works of visual art. Development of both systems commenced in the early 20th century and continues today. These systems define codes or notations that describe the content of an image and can assist in the search and retrieval of cultural heritage materials. Other classification systems (Baca et al., 2006, pp. 355-256) (Table 2.9) beyond those reviewed in this study.

Table 2.9 Sources for subject names and terms for Iconographic Themes

Classification system	Source
Illustrated Dictionary of Hindu Iconography.	Stutley, Margaret. London: Routledge and Kegan Paul, 1985.
Index of Jewish Art: Iconographical Index of Hebrew Illuminated Manuscripts.	Narkiss, Bezalel, et al. Jerusalem: Israel Academy of Sciences and Humanities; Paris: Institut de recherche et d’histoire des textes, 1976-1988.
Library of Congress Subject Headings.	Library of Congress. Authorities. Washington, DC: Library of Congress. http://authorities.loc.gov/.
Thesaurus iconographique: système descriptif des représentations.	Garnier, François. Paris: Léopard d’or, 1984.
Encyclopedia of Comparative Iconography: Themes Depicted in Works of Art.	Roberts, Helene E., ed. 2 vols. Chicago: Fitzroy Dearborn, 1998.

2.3.3.3.1. Iconclass

Iconclass is a classification system for iconograpic analysis and description of images or works of visual art: containing definitions of things that are, or may be represented or depicted. (Couprie, 1978; VanStraten, 1986). The product of more than 50 years work by Prof. Henri van de Waal in the study of western art, particularly (Becker, 1977; Trapp, 1976) and by a large group of scholars following his death in 1972, Iconclass was published between 1973 and 1985 by the Royal Netherlands Academy of Arts and Sciences (KNAW) and is managed by the Netherlands Institute for Art History in the Hague, since 2006. It comprises a classification system, an alphabetical index and a bibliography.

The classification system is a hierarchically ordered collection of 28,000 definitions of image content such as “objects, people, events and abstract ideas”. The alphabetic index contains 14,000 keywords to locate a notation and related textual description or index and the bibliography contains 40,000 references to books and articles of iconographical interest.

The hierarchical structure of the classification system is maintained by unique alphanumerical codes or *notations*. Ten primary divisions denoted by a one-digit *notation* may be further subdivided into a maximum of 9 subdivisions, by adding a digit to the right (Table 2.10). Subdivisions are further broken down by the addition of a *letter,* allowing up to 25 subdivisions (excluding J for legibility), extending the notation to describe or index subjects at a lower level of increasing particularity.

Table 2.10 Iconclass *notations* of top level divisions, general description and sub elements of notation 4[31]

Notation	Primary *Notations* (Top-level Divisions)	Type of descriptions	Example: Notation 4 sub divisions
0	Abstract, Non-representational		
1	Religion and Magic	1-5, general topics, designed to comprise all the principal aspects of what can be represented	41 material aspects of daily life
2	Nature		42 family, descendance
3	Human being, Man in general		43 recreation, amusement
4	Society, Civilization, Culture		44 state; law; political life
5	Abstract Ideas and Concepts		45 warfare; military affairs
6	History	6-8, special topics, coherent subject matter of a narrative nature	46 social and economic life, transport and communication
7	Bible		47 crafts and industries
8	Literature		48 art
9	Classical Mythology and Ancient History		49 education, science and learning

Text explaining the meaning of a *notation* is referred to as the 'textual correlate'. The ten primary/high level categories, the intended content type and an example, notation 4 ‘Society, Civilization, Culture’ contains the maximum nine second-level divisions (Table 6). Notation 42 ‘family, descendance’ is further divided into eight subdivisions 42A, 42B, 42C, 42D, 42E, 42F, 42G and 42Z. On the other hand, notation 6 ‘History’ contains only two second-level divisions: 61 ‘historical events and situations; historical persons’ and 62 ‘visions of the future’, whereas notation 4 ‘Human being, man in general’ has 4 second level categories.

Broader / narrower relationships are maintained through the notation coding system providing a tree structure, and the concept of *Auxiliaries:* bracketed text; keys; doubling of the letter; and structural digits, expands the functionality. Bracketed text allows a break away from the hierarchy of alphanumeric notations, for example, 98B(CAESER); keys contains a list of valid values that may be added to a notation to “add a shade of meaning”; doubling the capital letter highlights an opposite or emphasize an element (“25F Animals, 25FF fabulous animals, 31A the (nude) human figure [male]*), 31AA the (nude) human figure; female”.

[31] http://www.iconclass.org/help/outline Iconclass hierarchies online.

Iconclass is suitable for annotation of biblical and mythological subject matter (Hildebrand, van Ossenbruggen, Hardman, & Jacobs, 2009), evidenced by the depth of details regarding the apostles (Figure 2.2, Example 2). It may be contrasted with the richness of relationships and synonyms of the Getty thesauri. However, cross-reference search and discovery functionality reveals all occurrences (notations) and variations of a keyword or part-notation. Furthermore, the inclusion of bibliographic materials may provide historical and social contextual information to assist understanding and interpretation of art works.

Example 1: Search terms: **42F51** (maid) and **46E221** (letter, envelope) based on notations displayed when viewing the art work 'Lady writing a letter with her maid' by Johannes Vermeer (1632-1675)	**Example 2:** Search term **'Apostles',** notation selected: **73F21234**
4 Society, Civilization, Culture	7 Bible
42 family,descendance	73 New Testament
42F household management; housewife at work	73F lives and acts of the apostles of Christ; epistles
work	73F2 lives and acts of the apostles (part I)
42F51 maid ~ house personnel	73F21 life and acts of Peter
(civilization · culture · family · household offspring · society)	73F212 miracles of Peter (sometimes in co-operation with others)
46E221 letter,envelope.	73F2123 the high priest and Sadducees have Peter and the apostles put in prison (Acts 5:17-42)
	73F21234 the apostles are beaten, and are released afterwards ~ Peter and the apostles in prison
	Acts · Acts 05 · New Testament · Peter (St.) · Sadducees · apostle · beating ·bible · co-operation · high priest · miracle · prison · releasing

Figure 2.2 Example of Iconclass search terms

2.3.3.3.2. Index of Christian Art at Princeton University

The index was founded in 1917 by Charles Rufus Morey (1877-1955) at Princeton University in response to a growing interest in the subject matter in the department of Art & Archaeology (Hourihane, 2014) and based on the view that iconography could reflect cultural change in society. The purpose of the index is to provide an iconographic register of the subject of medieval art from the early Apostolic period up to the year 700, later extended to the end of the sixteenth century. The index comprises 2 files: a text file and a photographic file.

Text file: Over one million colour-coded catalog cards reside in the text file and are organized under 28,000 subject headings. Each card contains the same set of information about the artwork, including the title of the artwork; medium; location; date; provenance; bibliography; inventory number and a free-form description. Descriptions are objective and neutral and adhere to the same format and use the same terms, leading consistency of description across the index. Moreover, each artwork is assigned a primary subject and may have multiple sub-headings depending on the details depicted in the image. Scope notes, bibliography, and Iconclass notations completes the information documented about each artwork.

Photographic file: The photographic files contain 200,000 uniformly sized black and white images of the artworks and are filed under location within medium. Even though the images are black and white the colours are included in the descriptions (in the text files).

With the advent of technology the index was computerized in 1991 and work on the index continues as new artworks are discovered. The authors claim that that the Index is the largest and most specialized resource for the study of iconography in the world whereas Iconclass is the most widely used.

2.4. Web technologies for Cultural Heritage

The previous sections reveal evidence of considerable ongoing research into the standardization of technologies in the cultural heritage domain. Traditionally, institutions employ relational database (RDB) systems for collections management, with limited RDB content published online, through their custom designed web pages. Some museums also disseminate content to union catalogues. The potential for terminological standardization and knowledge sharing across heterogeneous data sets, from multiple sources, may be realized through the use of semantic web technologies. Currently, semantic web technologies utilizing Ontologies and Thesauri are deployed, for example in the biomedical, legal, geographic, cultural heritage domains. A review of recent research and web portals identifies critical aspects of implementing web technologies and highlights progress made in the cultural heritage arena. The section concludes with a discussion of how technology may contribute to research in the arts.

2.4.1. Research and Web portals

Common themes in the literature concerning semantic web technologies for cultural heritage include a) interdisciplinary collaboration b) computational ontology c) concept generation through data mapping from proprietary databases to conceptual models; d) knowledge base generation of collection instances, through data mapping from proprietary databases to standardized metadata schemata; e) semantic annotation of cultural content and f) search and retrieval. Also, essential issues concerning data quality, trustworthiness and provenance necessarily require consideration by content providers, aggregators and harvesters. These themes are explored through review of the ongoing research.

Evidence of the potential of semantic technologies enabling collection interoperability and intelligent applications in the cultural heritage domain was demonstrated more than ten years ago. The MuseumFinland semantic web portal (Hyvönen et al., 2005), which took second prize in the Semantic Web Challenge Award in 2004, represents an early exemplar and influential project. The study describes how seven ontologies along with object instances consolidated from four local collections databases into one repository, formed the foundation of a semantic view-based search engine. The resulting web portal presented multiple view groups of the classes of data (Artifact, Creation, Usage and Collection) and facets, organized as tree structures, presenting the different types of collection information and allowing further refinement of search criteria. The combination of views and facets assisted user search, retrieval and discovery of content and semantic associations across the ontologies.

With the increasing adoption of semantic technologies, the need for ease of access by users to cultural heritage repositories is paramount. Considerable research into semantic search and retrieval continues today, exemplified by a recent study facilitating more logically complex user queries than are currently supported (Isemann & Ahmad, 2014). Subsequent to the design (in cooperation with art experts) and population of an art ontology based on CIDOC CRM (CIDOC CRM: Conceptual Reference Model), this study investigates the use of Artfinder, a custom designed graphical user interface, whose search templates are generated from the underlying ontology and knowledge base. The system dynamically generates and processes complex queries, based on views, content selection and further search criteria refinements, presented in the interface. Results of the study involving 10 art specialists and 10 laypeople show that growing familiarity with the interface can assist users in expressing their intended queries and the authors propose that perhaps this semantically richer mode of accessing knowledge repositories will benefit domain experts and informed lay people (Ibid.).

Szekely et al. (2014) present the case of implementing semantic technologies for the Smithsonian American Art Museum (SAAM). The collaboration of computational ontology, database and art specialists implementing the technologies resulted in a web portal providing a view of the institution's 41,000 objects and 8,000 artists. The museum considered the Europeana Data Model (EDM: Europeana Data Model) appropriate as the underlying ontology because it "maximizes compatibility with a large number of existing Linked Open Datasets" (Ibid.). However, the project necessitated extending EDM to incorporate subclasses and properties to represent SAAM knowledge. Besides selection of an appropriate ontology, the authors identify three major steps in the implementation of a semantic web portal: a) mapping data from source/relational database technologies to semantic web technologies; b) linking data to external resources on the web and c) curating published data. The complex mapping process from TMS (The Museum System) relational database to RDF (Resource Description Framework) involved data preparation tasks of table filtering, data extraction and reformatting followed by allocation of data to the correct classes and properties and establishing relationships between the classes. Linked datasets included the Getty ULAN (Union List of Artists Names - p. 58), DBpedia [32] and data from the Rijksmuseum, providing supplementary details about the artists. Tombstone data extracted (and reformatted) from the existing web database comprised classification; role; descriptive terms and images for each artwork and name; title; biographical; geographical; birth and death date; residence and citation data for the artist (Szekely et al., 2013). In addition to achieving knowledge sharing and interoperability, SAAM ensured the quality and trustworthiness of published data whereby museum personnel verified all data generated and published and they recorded the source creator, comments and date updated for all data.

[32] http://wiki.dbpedia.org/

Further research focuses on the quality, interoperability and trustworthiness of published data (Georgiadis et al., 2014; Stathopoulou et al., 2014). Therein, the authors propose an infrastructure with separate dynamic components for validation of a repository at multiple levels: repository interoperability; metadata elements; metadata documents; and digital files, along with improving efficiencies for harvesting data from content providers. Accordingly, the resulting infrastructure separates validation rules for any level of a repository from the application code, thus enhancing long term maintainability. The authors report that they have developed a novel Validation Domain-Specific Language (VDSL) available through a graphical user interface, which empowers content providers to define arbitrary validation rules and logic as repositories and /or validation requirements evolve.

Adopting an alternative approach to SAAM, the Science and Technologies for Archaeological Research Centre (STARC) case study describes the online publication experience of both the Byzantine Museum and Art Gallery (Vassallo, Athanasiou, Hermon, & Eliades, 2013). These two Cypriot collections of the Archbishop Makarios III Foundation comprise paintings; frescoes; mosaics; statues and minor art objects (Ibid.) Selection and elaboration of the LIDO 33 model enabled standardized descriptions for the heterogeneous objects, at the required level of granularity, in both English and Greek. When converting the two original dissimilar databases to the customized LIDO schema, content was standardized and enriched by manual data insertion of: object ID; type; title; chronology; artist or school; provenance; material/technique; dimensions; conservation status; rights; donor; collection; place and so forth. This data enrichment provided comprehensive documentation and additional search support while simultaneously improving accessibility and usability of content through aggregation and publication to both the STARC repository and Europeana (EDM: Europeana Data Model)

Furthermore, implementing the LIDO schema for the STARC repository obviated complex mapping processes to contribute to Europeana, since Europeana is designed to ingest LIDO data.

Incorporating Linked Open Data (LOD) is becoming more common in semantic web portal design, given the advantages of acquiring supplementary information: for example, Getty ULAN contains information about artists that might not be commonly held in museum catalogues. Perhaps even more importantly for art specialist researchers, a recent study (Dijkshoorn, Aroyo, Schreiber, Wielemaker, & Jongma, 2014) moves beyond this issue of annotation: it seeks to determine if and to what extent LOD may broaden or diversify semantic search results. Base data for this study included collections data from the Rijksmuseum, museum query logs for a period of one month and five external vocabularies: Getty ULAN; Getty AAT, Iconclass, (Union List of Artists Names, Art & Architecture Thesaurus, Iconclass -refer section 2.3.2), WordNet [34] and IOC World Bird [35]. The authors matched search terms from query logs

[33] Lightweight Information Describing Objects http://www.lido-schema.org/schema/v1.0/lido-v1.0-specification.pdf
[34] https://wordnet.princeton.edu/ Last accessed 23 Nov 2015

with the collections repository alone, with each of the five vocabularies individually and then repeated the matching exercise five times with different dataset combinations, for example with the collection, AAT and Wordnet, the collection Iconclass and IOC and so forth, as deemed appropriate based on possible associations between the LOD sets. Perhaps not unexpectedly, results of the study show that incorporating external vocabularies increases the number of results: a simple example being that inclusion of Getty AAT yields 9 results for the term 'watercolour', instead of zero results, when searching the collection alone. More interestingly however, is the fact that, in some cases not only are more results found, but that the results are more diversified. For example, the term 'rembrandt' yielded additional clusters of data: paintings by Peter Lastman, a teacher of Rembrandt and paintings by Saloman Koninck, noted by ULAN to be an ardent follower of Rembrandt" (Ibid.) The authors contend that including LOD has the potential to satisfy the needs of humanities scholars or other groups, requiring results beyond the standard results. However, they advise that achieving such diversity of results is contingent upon the inclusion of rich LOD appropriate to the collection and the necessary existence of collections metadata elements to link to the LOD. Addressing some of the issues raised in that study, Chortaras, Drosopoulos, Kollia, and Simou (2014) also emphasize the importance of metadata, the connection of metadata to ontological domain knowledge and the need to satisfy requirements of experienced users such as researchers, professionals and other interested groups. These authors also argue that 'Semantic query answering' (search and retrieval) should be based not only on term matching on the collection but also on implicit knowledge found by reasoning over the terminological knowledge.

The case of intangible cultural heritage (ICH) is presented by Hu, Lv, and Zhang (2014). Chinese Folk customs, representing some of the most important ICH, are classified as "Worshipping and celebration rituals, temple fairs, traditional etiquette, festivals and apparel… and so forth" The Pang Wang Festival of Yao represents an exemplar test case from a vast array of intangible cultural heritage of the nation. The authors selected the CIDOC CRM ontology, as it defines the necessary events and properties to represent domain knowledge and the paper charts the process of constructing the ontology from the ground up, identifying the main classes: Activity; Time; Actor; Place and Thing, the kernel concepts of Class; Property; Domain; Range and Instance, and also defines the major properties and their definitions. Having constructed the ontology for the test case, the essential question of how to manage and maintain the ontology is addressed by storing every property in a database. Each property occupies a row in the database and all data required to reconstruct or update the ontology are stored as separate fields, in each property record. It is unclear from the paper if this test ontology was published. However, the process adopted by the authors establishes a method to digitally preserve, maintain and update intangible cultural heritage into the future.

[35] http://www.worldbirdnames.org/ Last accessed 23 Nov 2015

The Classical Art Research Online Services (CLAROS), an international interdisciplinary research project led by Oxford University (Kurtz et al., 2009) was officially launched in 2011, with the stated goal of bringing art to a global public and research continues as more content and institutions contribute to the site. Collaboration across disciplines and fields of classical archaeology; art history; museums; engineering science; and information and communication technologists results in a comprehensive resource discovery portal. The portal originally ingested heterogeneous content from five founding members, including: University of Oxford – Beazley Archive; University of Cologne – Research Sculpture Archive:Arachne; German Archaeological Institute; the University of Paris – Lexicon Iconographicum Mythologiae Classical; and University of Oxford – Lexicon of Greek Personnel Names. In-house systems at the five institutions comprise relational database technologies SQL Server, MySQL and other unnamed databases. CIDOC CRM forms the underlying ontology of CLAROS. Institutions contribute content by exporting a limited and common data set from their catalogue, applying CIDOC CRM concepts and expressing the data in RDF format. CLAROS provides comprehensive search and discovery features beyond the usual criteria, including: search by timeline; search by point on a world map; and search by image/part image, the former two options providing temporal and spatial contextualization of objects. Retrieved data is displayed and the user may click on it to return to the source provider's screen to display the full record.

Concluding this section on semantic web portals the case of the Europeana portal is presented. Europeana was developed under the Europeana Project, funded by the European commission with matching funding from Member and Nom-member states and operated by the Europeana Foundation based in the National Library of the Netherlands, in The Hague. The underlying ontology is EDM (the Europeana Data Model, Section 2.2.3.9) and was expressly designed for the Europeana project. Content for Europeana is harvested from content providers such as library, archives and museums, through National, Domain-level or Theme focused aggregators. Content control is managed through a Europeana Data Exchange Agreement between providers and Europeana. Providers export data from their databases as specified by Europeana (Europeana Project, 2014) and then converted to EDM (Ibid. Pp 35-45 documents a working example). Search choices and facets reveals the content harvested from providers. Indexing content is included in the initial search criteria presented to the user: All fields; Titles; Creators; Subjects; Dates/Periods and Places, to which the user may apply his or her search text to initiate a search. Additional services include the provision of Application Program Interfaces (API's) enabling end users to build applications using the functionalities of the Europeana Digital Library Management System (Concordia, Gradmann, & Siebinga, 2010). On a negative note, user query and retrieval operations highlight confusion between cultural material records and their associated digital records (Peroni et al., 2013), possibly stemming from data constraints in the original European Semantic Elements (ESE) model (Ref 2.2.3).

Alternatives to semantic web technologies in the cultural heritage domain include relational database systems, employed by museum systems worldwide. [36] The Giza Project at Harvard University combines traditional archives (dig photos, archaeological drawings, object metadata) with realistic 3D visualization of the site, providing a "revolutionary access to Giza" (Der Manuelian, 2013) and the authors claim that real-time immersive models allow one to pose new research questions, to provide interactive classroom instruction, and to investigate diachronic approaches to Giza's evolution over several millennia. The technologies underpinning the system include a SQL database.

The RKD Explore, a more recent innovation in the CH domain (Koot, 2015) presents a web portal for art historical research that incorporates 7 databases, comprising Artist, Image, Library, Collections, Portraits, Excerpts and the recently added Archives database. Database development at the RKD (The Netherlands Institute for Art History) began in 1990 with the RKD Artist database only and the author states that the technology the databases were built on remains largely the same.

In contrast to centralised repositories of ontologies and to LOD, to access Fine Art information, recent research investigates accessing this information through museum API's (Application Programming Interface), that provide search and retrieval of the museum's online catalogues (Buchanan et al., 2015). The authors propose that this alternative approach provides access to up-to-date information from trusted sources, through intelligent applications that automatically search participating institution's repositories, precluding the need to search individual website and for centralised repositories. For this researcher, the API approach suggests a pragmatic and promising solution for the researcher who requires the latest, reliable information resources, and especially given that museum systems increasingly adhere to domain metadata standards, facilitating interoperability, through API development and processing by intelligent applications.

By 2016, the future of the semantic web is not as clear cut as perhaps originally envisaged. Given the immense volume and diverse range of datasets on the web, Bernstein et al. (2016) assert that the challenge will be how to take advantage of these data for intelligent applications. Furthermore, they assert that methods that reflect how data is actually used, lie somewhere between the formal ontology approach at one extreme to unstructured and less controlled data at the other. They propose that research into "more meaningful services that rely less on logic-based approaches and more on evidence-based approaches", is required.

2.4.2. Discussion: Technology supporting Research in the Visual Arts

The literature shows that semantic web technologies are being investigated in the cultural heritage domain with projects such as Europeana and an Oxford University led project CLAROS, available on the web. When researching a work, art experts rely on trusted sources,

[36] http://collectionstrust.org.uk/software/ last visited 25/02/2018

and up to date information. Regarding centralised repositories, such as Europeana one accepts the authenticity and trustworthiness of the data, since Europeana and its' authorised aggregators control data ingestion. The question arises, however about the frequency of data ingestion or more precisely, if data is amended at the source after it has been uploaded to Europeana, by way of example, if provenance research results in changes to the attributed artist can and when is that change reflected in the centralised repository.

Web portals serve to promote collections and their respective institutions to a global audience. The degree to which technology supports art specialists conducting research in the visual arts is less clear, however. Recent research shows an increasing interest in artist's archives as potentially valuable resources, providing background information about the artist and the context of the creation of a work of art, but there is little evidence of implementations that incorporate these materials with a work of art. An exception to this is the recent addition of an Archives database to the web portal by the RKD. The value of artist's archives as rich resources for art research warrants further investigation and is particularly relevant to this study that aims to incorporate the knowledge of art experts in a representation that may support research in art.

As discussed earlier, recent research suggests that web technologies may support scholarly research processes in the arts and it addresses some of the stated needs of these researchers. That is, work has advanced on the standardization of metadata schemata and ontology models, facilitating an agreed representation of collections data and agreed terminology of the domain, respectively. Furthermore, these technologies enable interoperability across technologies and collection repositories where thesauri can not only enhance intellectual access to art objects through semantic annotation of terms but also through defined relationships enabling linkage of an object to another set of objects. Semantic search and retrieval systems also assist users in specifying search queries to locate images and data.

It appears that in these web portals that content is presented in read-only mode and that all of the content pertaining to the artefacts presented, originate from systems in each of the contributing museums. Arguably, the extent of the content and quality of the content presented on the web is dependent upon these original data sources. The quality of the original sources, however, relies upon adoption of the appropriate metadata schema for the different types of collections and the consistent application of the metadata to ensure that content is always stored in the correct metadata element and in the same element for all works in the particular collection. In addition to the above conditions, good cataloguing that captures content beyond the common tombstone data will provide the kinds of details that may support domain expert research.

Finding from recent studies identify shortfalls that may hinder art research through the use of technology, including: a) missing online art collections and artists archives, b) missing and/or lack of high quality images online, c) the need for more subject terms/search keywords for art objects and d) need for improved descriptions of art objects.

2.5. Summary

The literature reviewed encompasses the theoretical foundations and practical research underpinning this study. A review of knowledge representation literature identifies principle rules for implementing a representation and the characteristics of a good representation of the real world. Research into ontology as adapted by the discipline of computer science demonstrates some influences of Ontology from its philosophical origins, through the organised classification of the kinds of things such as substances and abstract entities and their relationships and general categories that may apply across domains. An alternate technology for representation of the real world includes the long established relational database, whose strengths include efficient storage and retrieval of large volumes of data and the multiple relationships that may be established between the entities represented in the data. The complexity of aesthetic objects is revealed through a review of the literature on the ontology of art and this along with art specialist practices must be incorporated in a computational system whose aim is to represent the ontological commitment of art specialists and aimed at supporting research in the arts. The literature also revealed the value of artist's archives for art historical research and these must also be considered in this study. A considerable body of knowledge in the cultural heritage domain is evidenced by the increasing sophistication of metadata schemata, supporting data consistency, resource discovery and interoperability. Also, a review of these knowledge objects such as thesauri or structured terminologies and classification systems uncovers the immense body of expert knowledge developed over decades by multidisciplinary teams invested in the preservation of cultural heritage. Finally, the practical research presented concerning web portals provides evidence of the continual progress made in the representation of cultural heritage.

3. Methods

This chapter describes the theory and practical methods employed in this study, towards fulfilling the primary aim which is to incorporate the knowledge of domain experts in a computational representation that describes important aspects of the domain and that may serve as the basis for applications supporting scholarly research in the domain.

Achieving that aim requires an understanding of the role of the curator and the kinds of things these domain specialists consider important when researching a work of art. Since this researcher has limited knowledge of art, and in that sense would have negligible bias regarding the domain and domain methods, it was clear from the outset that access to the expertise of domain specialist was needed. To that end, and having considered other art institutions, we approached experts at the National Gallery of Ireland (NGI).

Consequently, this research is carried out in cooperation with experts at the NGI, in that, it is informed by inputs from curators, an archivist, a collections management system administrator and a research curator, over many interactions with these experts. Additional inputs from multiple sources provided deeper insight into important aspects of the domain from complementary perspectives, including a survey of art experts based in different museums, in several different countries.

Decisions about the research approach, influenced largely by the worldview of the researcher and the nature of the enquiry, determine the research approach, design and methods used in the enquiry (3.1). Methods of conceptual modeling are fundamental to the representation of the domain and future system design (3.2). Furthermore, a representation grounded in mathematical theory underpins this study (3.3). Principles for robust system design and implementation based on the relational model are considered (3.4). The steps involved in designing a relational database are discussed (3.5). Methods of text analysis and harvesting archives into a relational database (3.6) provided the external contextual information for enhanced representation of aesthetic objects, central to this research. The next section describes data collection methods and analysis techniques for knowledge acquisition and refinement of research question (3.6). The chapter concludes with a discussion of the methods employed in this study (3.7).

3.1. Research approach

The selection of a research approach is not only informed by the nature of the research and the research question but also by the philosophical worldview assumptions of the researcher regarding the two main philosophical dimensions of a) the perception of reality (ontology) and b) the investigation and development of knowledge (epistemology) (Creswell, 2014; Venkatesh et al., 2013; Venkatesh, Brown, & Sullivan, 2016; Wahyuni, 2012). In turn, the researcher worldview influences the research design, methods and instruments for data collection and analysis. Different worldviews or “beliefs that guide action” (Creswell, 2014) are reflected in the research methodology selected by the researcher.

3.1.1. Research Methodologies

Two main research approaches, *qualitative* and *quantitative* are identified in the literature (Bryman, 2012, p. 116; Creswell, 2014, pp. 5-10; Kothari, 2004) However, Creswell identifies a third, and more recent approach, *mixed methods research* that employs both qualitative and quantitative methods. Each of these approaches encompasses different data collection methods and analysis techniques and determines how data are interpreted.

Quantitative research design involves data collection and analyses based on experiment, survey, or structured interviews to test a theory or hypotheses. This type of design investigates the measurement of relationships between variables: cause and effect, correlation, trends of variables in a population and this approach reflects a positivist philosophy, based on systematic observation and computation of the objective reality in the world.

Collection and analysis of data from multiple sources, to arrive at a more holistic and complex account of the issue under enquiry, rather than relying on one data source, underpin *Qualitative* research (Creswell, 2014, pp. 183-189). Moreover, as the key instrument for data collection the researcher collects the data in its natural setting and seeks to understand the meaning that participants hold rather than any meaning from the researcher’s perspective, and thus this research reflects an interpretivist or constructivist worldview.

A relatively new design, *mixed methods* research (MMR) strongly emerged in the 1980’s. This approach involves collecting and analyzing both qualitative and quantitative data, the combination of which provides richer insights and deeper understanding of the problem area, rather than using one method alone (Creswell, 2014, pp. 215-219; Johnson & Onwuegbuzie, 2004; Venkatesh et al., 2016) while overcoming the bias and weaknesses of each form of data. This reflects a more pragmatic philosophy where objectivist and subjectivist views of the world are not mutually exclusive. Rather, pragmatism arises out of actions, situations and consequences and the researcher considers all approaches available to better understand the problem.

This study incorporates the knowledge of museum curators responsible for the preservation of collections their care and it necessarily involves qualitative methods, by way of example, interviews, meetings and observations, for this researcher to gain insight into important aspects of the domain from the art specialist’s perspective. Thereafter, knowledge of the wider community of domain specialist’s conceptualization of art and of their information needs offers a complementary view of the data and tests the prevalence of the qualitative views among the wider community. Attaining this knowledge requires survey methods. Findings from this phase of the study inform the design of a computational representation of aspects of the domain.

Given the nature and procedures of this study, a mixed methods approach offers the best design approach.

3.1.2. Mixed Methods design

While a mixed methods research approach (MMR) enables stronger evidence (Creswell, 2014, p. 4) through complementary (both qualitative and quantitative) methods of data collection and analysis, both of which are based on the same questions, MMR may be more difficult to execute (Yin, 2014). Consideration of three predominant MMR models (Creswell, 2014, pp. 219-229; Venkatesh et al., 2016) assists with the selection of an appropriate model for this study:

a) *Convergent parallel mixed methods*: Collection of both qualitative and quantitative data provides a comprehensive analysis of the problem. Both datasets test the same or parallel variables and constructs and the researcher analyses both datasets separately. Comparison of the datasets reveals convergence or divergence of data and findings are based on the integration and comparison of both sets of data.

b) *Explanatory sequential mixed methods*: In this two-phase approach the researcher first collects and analyses quantitative data and then selects aspects of the findings for further exploration and elaboration through qualitative methods.

c) *Exploratory sequential mixed methods*: The reverse of b) above, the findings from an exploration and analysis with qualitative data feed into a quantitative phase seeking to measure the extent to which findings from the qualitative phase can be generalized to a larger sample. Also, findings from the first phase may inform subsequent research, perhaps refining research sub-questions and or identifying appropriate instruments to use and appropriate variables to include in the quantitative research phase.

Initial readings of these models suggest that c), an *Exploratory sequential mixed methods* model is appropriate for this study. The qualitative phase would facilitate Interviews, observations and fieldwork to gain insight into the domain and inform design of a survey instrument for the subsequent quantitative phase that would provide a complementary perspective of the domain, while also testing the prevalence of finding from the qualitative phase over a wider population.

In addition to adopting one of the above mixed method models, the model may represent a specific phase in more complex or advanced research strategies, such as transformative, multiphase and embedded approaches (Creswell, 2014, pp. 227-239; Yin, 2014, p. 193).

a) A *Transformative mixed methods* strategy applies for example to areas of social justice dealing with people experiencing discrimination or oppression.

Since this study does not seek to explore and /or transform aspects of society or societal problems, this strategy is not appropriate.

b) In a *multiphase mixed methods* design, any of the models may be used together with other research methods in longitudinal studies for evaluation long-term program goals.

   Again, and although this study proceeds in multiple sequential phases it does not seek to measure or evaluate long term goals and therefore this strategy is also rejected.

c) In an *embedded mixed methods design* any of the above models may represent a supporting phase within a larger research design, for example, it may inform the design of an artifact or experiment.

The purpose of this mixed methods study is to gain insight into the domain of fine art and the knowledge and information needs of art specialists responsible for the preservation of collections in their care. This knowledge will be incorporated into the representation of the objects and collateral materials from primary sources of artist archives, but the results of the MMR do not dictate the design and functionality of applications underpinned by that representation. Therefore, the initial exploratory sequential mixed methods approach suffices for this thesis.

Consideration of both the qualitative and quantitative data collection methods and analysis techniques continues in 3.7.

3.1.3. Ethics

Defined as “The codes of conduct or moral principles recognized in a particular profession, sphere of activity, relationship, or other context or aspect of human life” (OED), ethics is fundamental to the integrity of academic research. Accordingly, ethical considerations need to be addressed at all key stages of research, including, prior to a study; study commencement; data collection; data analysis; and dissemination (Creswell, 2014, pp. 92-101). Central to much research, specifically in social science and education research and to all studies involving people, is the appropriate treatment of human subjects (Creswell, 2014; Yin, 2014, pp. 77-78). Silverman (2011, pp. 87-110) emphasizes the need to be honest and fair to participants by explaining the true purpose of the research with no attempt to conceal, deceive or exploit the participants. Furthermore, the researcher must ensure that the source of information remains private where confidentiality has been promised. Increasing awareness of ethical principles in the area of research in information technologies (King, 2015; Myers & Venable, 2014; Stahl, Eden, Jirotka, & Coeckelbergh, 2014) builds upon established ethical practices in research to considerations of consent, quality of artefacts and to broader issues such as information security and the benefits of technology to society. Meanwhile, work continues on updating a code of ethics for responsible computing in the profession (Brinkman et al., 2017).

This research complies with ethical standards of the School of Computer Science and Statistics, Trinity College Dublin and required review and approval by the Research Ethics Committee. The research application to the committee included a brief proposal and outlined the purpose of the study along with the research methods and instruments, types of participants, data protection and legal requirements involved. Included also were the required Information Form and Informed Consent Form for study participants (Appendix F).

This research is carried out in cooperation with domain experts at the National Gallery of Ireland and selection of a purposive, representative sample of participants (Kothari, 2004, pp. 51-59) comprised domain experts from NGI such as curators and archivists and a curator from the Chester Beatty Library. Participants received the Information form and Informed Consent forms in advance of pre-arranged interviews and the researcher retains signed copies of the consent forms from interviewees. Both of these forms were also required to be presented for verification and acceptance by survey participants on the initial pages of a web based survey. Furthermore, the Board specified that all survey questions must be optional and participants were not compelled to respond to a question.

Data privacy and security measures: All data collected were stored in secure data stores. Transcriptions of audio interviews were placed into password protected documents and survey data was collected directly into a custom built, private and secure database. Data processing and analysis details reside in password protected spreadsheets.

3.1.3.1. Copyright

Prior to publication of the web survey the researcher purchased licenses from the Irish Visual Artists Rights organization (IVARO) for images of two artworks by artist Jack B Yeats, as these are still protected by copyright. On expiry of the licenses (after a period of six months) the researcher replaced the images with open access images from the National Gallery of Art, Washington.

3.2. A Conceptual Data Model

Representing aspects of reality (universe of discourse UoD) in a computer system requires an understanding of the kinds of things of interest to stakeholders of the system and the relationships between those things. Recall (2.1) that the correspondence between the machine representation or ‘surrogate’, rests upon the quality of the mapping between it, ‘the representing world’ and ‘the represented’ or real world. Winston (1992, p. 18) characterizes a good representation as one that explicitly defines the important objects and their relationships in a concise yet complete manner, suppressing irrelevant details until required. Moreover, Winston states that good representations are computable and enable rapid storage and retrieval of information.

Modelling the real world has been addressed by database experts over many decades, where they abstracted important data using methods of conceptual modelling. Accordingly, as the first step towards a good representation of the UoD, the researcher designed a conceptual model as this provides the methodology for representation in terms of the real world objects and their relationships and is fundamental to successful system design (Coronel & Morris, 2016, p. 169; Elmasri & Navathe, 2014, p. 201; Elmasri, Weeldreyer, & Hevner, 1985; Kroenke & Auer, 2012, p. 156; Teorey et al., 1986). Used during the top-down requirements analysis phase of system design, the outcome of the data modeling process consisted of a high level, conceptual model of the world, depicted in diagrammatic form. Representing an abstraction of the objects of interest and their relationships the model also documents the basis of information requirements and future applications development. Tools and techniques employed in this modeling process included the Entity-relationship data model and aspects of the Extended or Enhanced Entity-Relationship (EER) to capture more complex requirements and semantics of the domain.

3.2.1. Entity-Relationship (ER) model

The entity-relationship (ER) data model, including a graphical technique for conceptual and logical representation of data was introduced in the 1970’s (Chen, 1976). Chen proposed the ER data model as a unification of previous data models, including the network model (Bachman, 1969), the relational model (Codd, 1970) and the entity set model, by incorporating data independence and semantic information presenting a more natural view of the real world consisting of entities and their relationships.

Central to Chen’s ER model are three kinds of objects: entities, attributes and relationships. Specific symbols identify each of the different objects in the model (Figure 3.1)

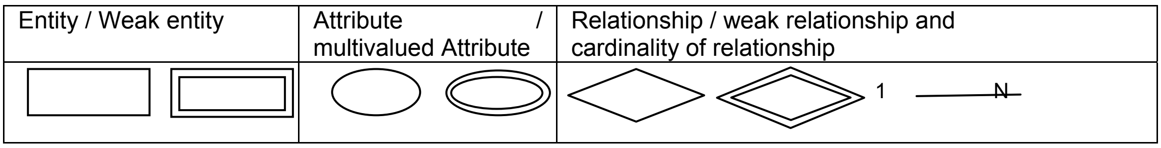


Figure 3.1 Entity-relationship model symbols

a) *Entity (and Entity Set).* Entities represent instances of both physical and intangible objects that exist in the world and in our minds (for example, an Artist or an Artwork, Provenance of a painting and an Exhibition). Each entity may be classified into a different entity type or entity set such as Agent, Work and Label (Figure 3.2) and all entities in an entity set have common properties. For example, the entity set Agent/Artist contains (among others) the artist entity Jack B Yeats and entity Johannes Vermeer.

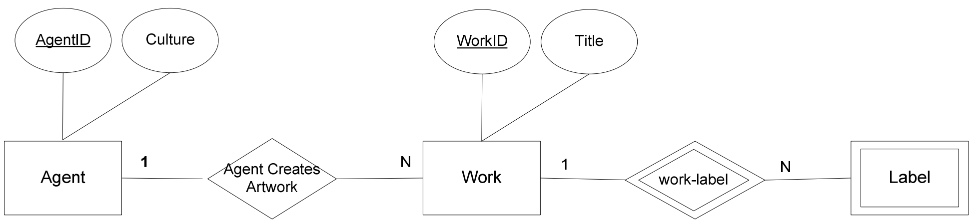


Figure 3.2 Sample ERD depicting three entities.

An Agent (entity) creates Works (entities). Curators may write many different labels (entities) for an artwork. Since a Label depends on the artwork for its existence and identity, labels are characterized as *weak* entities.

b) *Attributes (Value and Value Set).* Attributes refer to the properties of interest about an entity or a relationship, described by a set of attribute-value pairs. Value sets (domains) restrict the types of values that may be mapped to an attribute value. By way of example, the value sets Colour, FirstName, LastName and Integer determine the types of values, such that the values red, green, blue belong to the value set Colour; Jack and Johannes belong to the FirstName value set while Yeats and Vermeer would belong to the Last Name value set and 10.2 belongs in the numeric value set. Although attributes and value sets may have the same name, they are different concepts and the attribute Name may map to value sets FirstName and LastName, Width and Height of a painting may both map to the numeric value set.

Besides describing an entity, an attribute or a combination of attributes uniquely identifies an entity. If one or more sets of attributes can uniquely distinguish between entities (known as candidate keys) then one is arbitrarily selected as the unique identifier, known as the Primary Key.

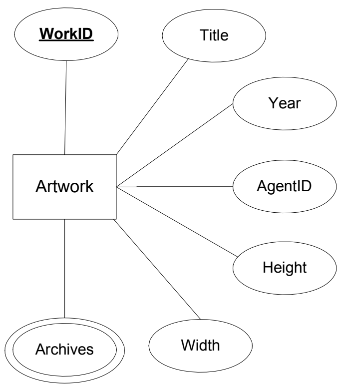


Figure 3.3 Attributes (oval shapes) for Artwork entity

Unique identifier / Primary Key is ArtworkID (underlined) and other attributes describe the entity.

If multiple attributes are required to uniquely identify an entity then the designer may choose to assign a surrogate key, such that the value of the surrogate key maps to a value set that ensures uniqueness, for example, the set of integers that automatically increment for each entity. The archives attribute represents multiple attributes perhaps including different sketches, diary entries, photographs, letters and other types of archival materials.

c) *Relationship, Role and Relationship Set.* A relationship is an association among entities, and a relationship set is a set of relationships of the same type. Chen states that a *relationship set Ri* is a mathematical relation among *n* entities, each taken from an entity set:

$\{[e_1, e_2, e_3, ..., En] \mid e_1 \in E_1, e_2 \in E_2, ... , e_n \in E_n, \quad \}$

and each tuple of entities [e_1, e_2, e_3, …., e_n] is a relationship.

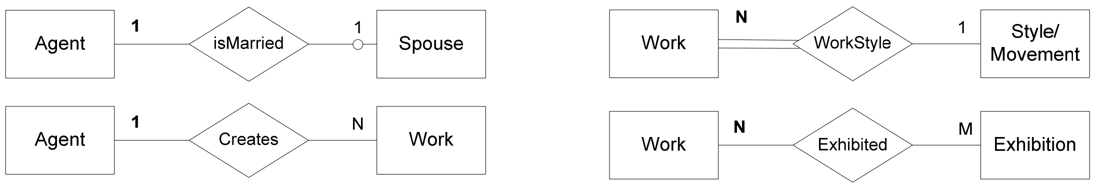


Figure 3.4 Relationships among entities.

1, N, M denote one or more entity occurrences involved in the relationship. All relationships are mandatory except Agent isMarried, denoted by a circle.

Relationships of interest to the curator in an art gallery may be those between artworks exhibited in exhibitions or the provenance of an artwork detailing movements between dealers and sellers or purchasers over the course of time. Alternatively, an attribute or attributes of an entity may also be used to form relationships. By way of example, connecting the ArtistID attribute from the Artist entity type to the ArtWorkID attribute in the Artwork entity type, represents the relationship described as 'artist creates an artwork'. Similarly, connecting the attribute StyleMID from the StyleMove entity type to the attribute aStyleMID in the Artwork entity type denotes the relationship 'many artworks may belong to a particular artistic style or movement'. Furthermore, the entities in a relationship need not necessarily be distinct. By way of example, a recursive relationship exists between a Series entity in the archives entity set and one or more Sub-Series in the same set and in that case the Series serves the role of parent, and the sub-series serves the role of child of the series.

Semantic information pertaining to relationships (Figure 3.4) comprises the following constraints. The degree of a relationship means the number of entity sets involved in a relationship, where an n-ary relationship is of degree n, and unary, binary and ternary relationships represent 1,2 and 3 entity sets respectively. The connectivity of a relationship describes the mapping of associated entity occurrences in the relationship, which can be 1 or Many; the cardinality ratio (the actual number of the entity occurrences in the 'many' side in a relationship may be shown in upper-lower bounds; and the participation membership class denotes whether an entity's participation in a relationship is either mandatory or optional. Double lines indicate that all entities in the set participate in the relationship.

Attributes may also be defined on relationships between two entity sets. Consider for example that the system needs to track the different locations associated with an artist. Attributes such as FromYear and ToYear and the association such as 'lived in', 'visited' or 'studied at' may be used to describe the relationship, but none of these attributes are defined on either the Artist or the Location entity. Instead, these attributes may be mapped to a relationship relation such as Artist-Location.

Chen characterizes entities and relationships as either *regular* or *weak.* Regular entities are those things that can be uniquely identified by a key, such as an artist or a painting. *Weak entities* on the other hand, are instances whose existence depends upon their relationship to another entity, and can only be identified through that relationship. By way of example, a curator at the gallery designs a label describing an artwork and over time the curator amends the label for exhibition purposes or perhaps new information about the artwork comes to light. Entity type 'label' may be classed as a weak entity (Figure 3.2). A regular relationship entity is one in which each participating entity is uniquely identified by its own attribute values or primary key, whereas a *weak relationship* relation is one in which a participating entity itself relies on another relationship.

3.2.2. Enhanced Entity-Relationship (EER) model

With the increasing complexity of applications and information requirements, limitations in the original ER model to accurately represent the reality of the domain, emerged. Continual research into more precise representation resulted in the enhanced entity relationship model (EER). In addition to the fundamental concepts of entity, attribute and relationship, the EER contains constructs such as subset hierarchy, generalization hierarchy, category and aggregation. (Elmasri & Navathe, 2014; Elmasri et al., 1985, pp. 201-286; Teorey et al., 1986). EER supports a more accurate representation of reality and the data requirements because "*it requires one to focus on greater semantic detail in the data relationships*" (Teorey et al., 1986). Consideration of properties of the EER follows.

a) Subset Hierarchy and Generalization

In a subset hierarchy, an entity E_2 is a subset of E_1 if all instances of E_2 also occur in E_1, denoted by $E_1 \subseteq E_2$. In this case every instance in the generic entity E_1 may also occur in one or more possibly overlapping subsets of entities (Figure 3.5b) reflecting the distinct role of the entity in each specialized subset. In a generalization hierarchy an entity E is a generalization of a set of entities E_1, E_2, E_3,..E_N if each instance of E is also an instance of one and only one of the set of entities E_1, E_2, E_3,..$E_{N.}$ (Figure 3.5c). In this case subclasses are disjoint.

Classifying objects into superclasses and subclasses is achieved through either a generalization process or specialization process. The Specialization process is a top-down process of conceptual refinement, first identifying a general entity type and then successively defining subclasses for more specific groups of entities (Elmasri & Navathe, 2014, pp. 246-259). The inverse process, Generalization, is a bottom-up process of conceptual syntheseis, commencing with identifying specific entities (where painter, sculptor, engraver and photographer may be relevant to this research) and generalizing these to a superclass concept, such as an Artist superclass (Ibid.). Attributes of the generic or superclass comprise only those characteristics common to all of its subclasses. Subclass/superclass relationships represent IS-A (IS-An instance of) relationship, for example, Painting or Sculpture IS-An artwork, Painter IS-An Artist.

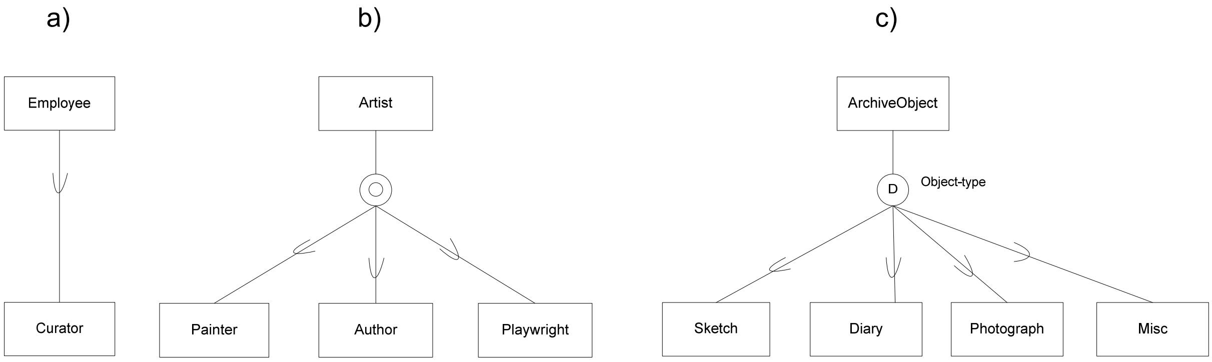


Figure 3.5 Subclass/superclass relationships

a) - Curator entity is a subclass (specialization) of superclass (entity) employee.

b) - Members of the Artist superclass may be a member of one or more (overlapping) subclasses, representing possible different roles played by an artist - Painter, Author and Playwright

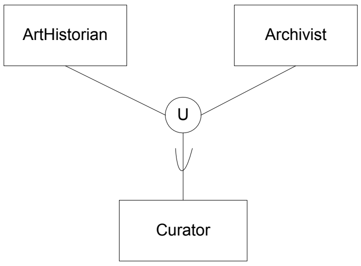
c) – Members of ArchiveObject superclass occur in only one of the (disjoint) subclasses Sketch, Diary, Photograph or Misc Ephemera

Membership and constraints on superclass/superclass relationships: Members of a subclass are always members of its superclass and they represent the same real world entity as that in the superclass. However, while members in *overlapping* subsets represent the same real world object defined in each (possibly) overlapping subclass (albeit in a different role) members in *disjoint* subclasses represent distinct real world objects. Moreover, a single attribute in the superclass (known as the 'defining attribute' or 'subtype discriminator') distinguishes between subclasses into which an entity is placed. By way of example, the attribute object-type in the ArchiveObject superclass determines the appropriate subclass of an entity (Figure 3.5).

Alternatively, multiple attributes may also be used for assignment to subclasses, for example, Boolean attributes IsPainterYN, IsAuthorYN and IsPlaywrightYN which the user selects when creating or modifying an entity (Figure 3.5b). Also, members of a subclass inherit all properties (attributes and relationships) defined on the superclass and additional properties may be defined on a subclass, the benefits of which include minimal occurrences of null attributes in the generalization / superclass and particular relationships need to be defined for specific subclasses only. Allied to disjoint and overlap constraints is the Completeness constraint. Total completeness means that every occurrence in the superclass must be a member of at least one subclass, while partial completeness means that not every member of the superclass needs to be a member of a subclass.

b) Categories

The EER model also includes the new concept of Category (Figure 3.6). In the ER model each entity type contains common entities and the entity type plays a particular role when participating in a relationship. Where more complex operations and relationships exist however, sometimes it may be necessary to define a single class that represents a collection of entities from *different entity types*. The ‘category’ construct represents a subclass of the UNION of members from two or more distinct entity types. Members of the category must exist in one of the participating entity types and each entity in the category subclass inherits the attributes of the superclass to which it belongs. The advantage of a category is that the category may participate in any relationships, thereby allowing different entities to participate in the same role in the same relationship.



A category may be classified as ‘total’ or ‘partial’. A total category is one in which the union includes all entities from each participating entity type, whereas a partial category holds a subset of the union of the entity types.

Figure 3.6 Curator Category - The union of ArtHistorian and Archivist entity sets.

c) Aggregation

Aggregation represents a relationship between a whole object and its component parts (Elmasri & Navathe, 2014, pp. 228-230, 270-271) and the concpet of aggregation applies to the EER model in three ways. The first case refers to the aggregation of a collection of attribute values to form the whole object. The second case refers to representing an aggregate relationship as an ordinary relationship. The third case, although not represented explicitely in the EER, combines objects already related by a particular relationship instance into a higher level aggregate object. The higher level object itself may be related to another object, where the relationship between the primitive objects and their aggregate object represents an Is-Part-Of relationship and the inverse is called Is-A-Component-Of (Ibid.) J. M. Smith and Smith (1977) propose aggregation as an abstraction of relationships between objects into a higher level object. By way of example, the entities and relationships involved when a curator plans an exhibition of paintings, presented in a specific location for a particular time span, may be abstracted as an aggregate object ‘exhibition’. The concept ‘exhibition’ may then be used without referring to all of the underlying details of the multiple entities and relationships involved.

3.2.3. Conceptual model for this research

Based on preliminary explorations of the domain an initial conceptual model (Figure 3.7) for the proposed system presents the important entities and the relationships between these objects.

The model proposes an enhanced representation of art that includes primary sources of archival materials, providing contextual details of interest to domain experts involved in art research. Entity types Work (for Artwork) and Archives represent superclass / subclass relationships of disjoint specializations. The category ‘Exhibit’ enables one or both of artwork entities and/or archive materials to be shown together in multiple exhibitions. Exhibition and

Provenance may be viewed as aggregations since location, people, temporal information and other associated details are omitted from the model.

Identifiers (primary keys) and attributes are omitted from this high-level conceptual model and the cardinality of relationships remains as many-to-many and will be resolved during subsequent decomposition into a logical model, aligned with a particular persistent storage model.

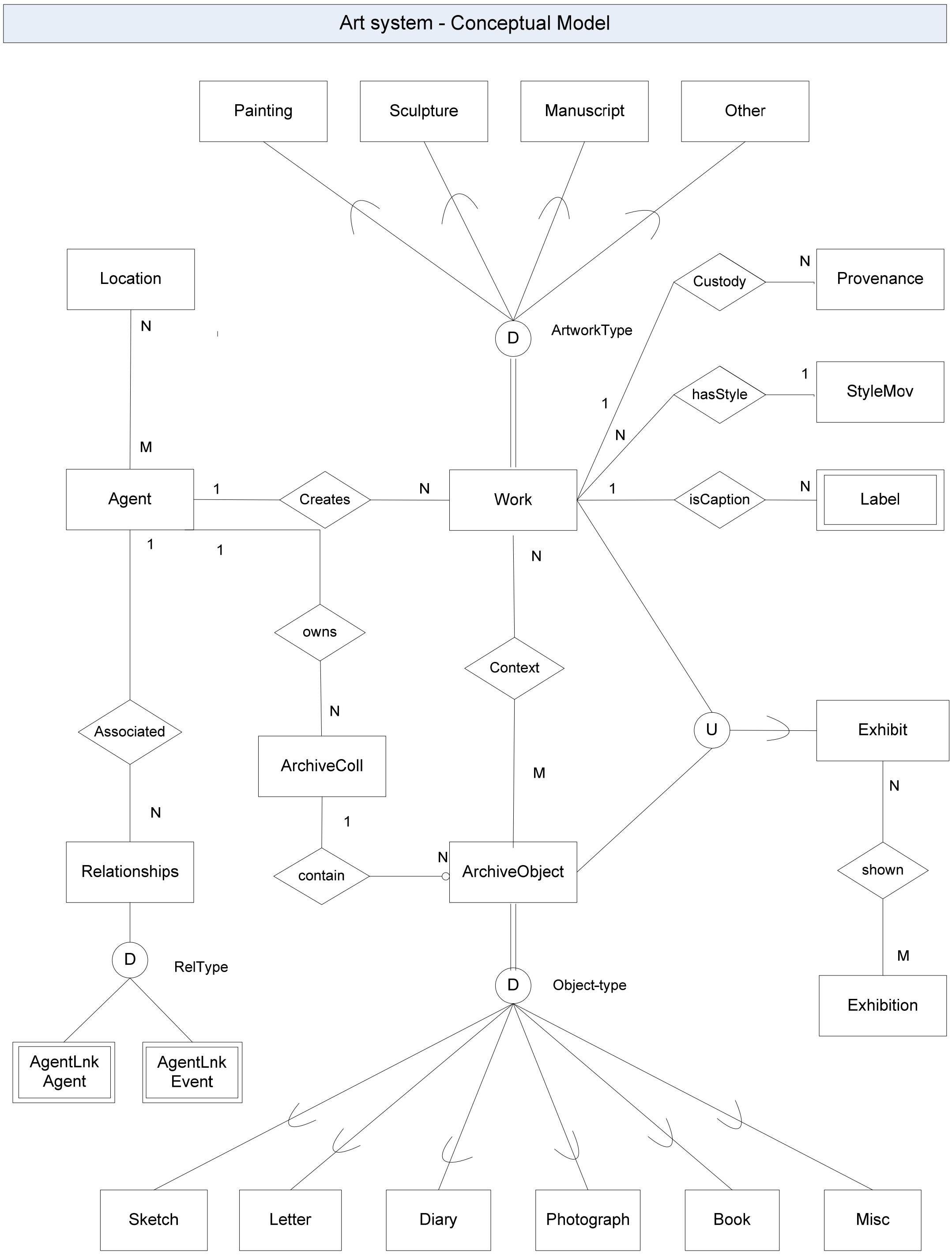


Figure 3.7 Conceptual Model – Enhanced Entity Relationship Diagram (EERD)

3.3. Technologies for persistent representation

In this section I will rehearse the fundamentals of relational technology in some detail. This section can be skipped in the first reading.

3.3.1. Notes on Relational Model Theory

In practice, data in the relational model are presented in tables and the practical language of the relational model is SQL (Structured Query Language). However, the theoretical foundation of data organization and operation is grounded in mathematical Set theory (Codd, 1970) and "*Since relations are sets, all of the usual operations are applicable to them*", in addition to operations specifically designed for the relational model. Further emphasizing the significance of the mathematical basis for the relational model Codd (1979) asserts that the basis for representation and subsequent manipulation by applications rests not only upon the structures (domains, tuples and relations) but in a large part upon the algebraic operators as enablers of multiple different views of the data. Moreover, Codd notes the close relationship between the relational model and first-order predicate logic. Distinguishing the relational model from other data models Codd suggests that a "*fully relational*" model supports a) the structures of the relational model, b) specified insert-update-delete rules and c) a sublanguage that is at least as powerful as the relational algebra. Consequently, in addition to the procedural language of relational algebra, relational calculus provides a formal higher level declarative language to query relations (Codd, 1971; Elmasri & Navathe, 2014, pp. 148-200).

3.3.1.1. Set Theory operations

Set theory operations comprising Union, Intersection, Minus, Cartesian Product and Division (Table 3.1 Relational Algebra Operations from Mathematical Set theory.) apply to two or more (binary or n-ary) sets of tuples, in the relational model. UNION operations combine all tuples from the original relations, by appending all tuples into a result relation (but discarding duplicates from the result set). The INTERSECTION operation results in a new relation comprising only tuples that are common to both relations, whereas the DIFFERENCE (also known as MINUS) operation results in a relation that contains only those tuples from the first relation that are not in the second relation. Adapting Set theory operations to the relational model, requires that the relations involved in UNION, INTERSECTION & DIFFERENCE operations must also satisfy *union compatibility* or *type compatibility* constraints, such that, relations have the same degree (number of attributes) and that the domain of corresponding attributes from all tuples are type compatible. However, the attribute names in both relations may be different.

In contrast to Union, Intersection and Difference, where the result relation has the same degree as the input relations and the number of tuples is always <= R_1 + R_2 (duplicates being discarded), the CARTESIAN PRODUCT (also known as CROSS PRODUCT) results in a relation where each tuple is a combination of all attributes from the first relation followed by all attributes of the second relation, therefore type compatibility is unnecessary for this operation.

The number of tuples in the result equals the number of tuples in the first relation multiplied by the number of tuples in the second relation (R_1 X R_2.)

Finally, the DIVISION operation is used "*to answer questions about one set of data being associated with all values of data in another set of data*" (Coronel & Morris, 2016, p. 90), resulting in a set of tuples from the original relation (treated as the dividend) that match all tuples in the dataset used in the second or search relation (treated as the divisor).

Table 3.1 Relational Algebra Operations from Mathematical Set theory.

Adapted from Elmasri & Navathe (2014)

Operation	**Type**	**Notation**	**Produces a Relation that includes:**
UNION	Binary n-ary	$R1 \cup R2$ = $R_3 \cup R_2$ $R_1 \cup (R_2 \cup R_3)$ = $R_1 \cup R_2) \cup R_3$	Includes all tuples in all of the original R_1 or R_2 … Rn relations, but duplicates are excluded. Equivalent SQL clause is UNION
INTERSECTION	Binary n-ary	$R1 \cap R2$ = $R_2 \cap R_1$ $R_1 \cap (R_2 \cap R_3)$ = $R_1 \cap R_2) \cap R_3$	Includes all tuples that are common to all of the original relations. Equivalent SQL clause is INTERSECT
MINUS (DIFFERENCE or EXCEPT)	Binary	$R_1 - R_2$ Note: $R_1 - R_2 \neq R_2 - R_1$	Includes all tuples from R_1 that are not in R_2 Equivalent SQL clause is EXCEPT
CARTESIAN PRODUCT (also known as **CROSS PRODUCT)**	Binary n-ary	R_1 x R_2	Includes all attributes from R_1 and R_2 (Degree of new relation = degree R_1 + degree R_2 AND the resulting relation includes all possible combinations of tuples from both relations, R_1 x R_2 * type compatible not required. Typically useful when followed by a SELECT clause.
DIVISION	Binary	$R_1 (Z) \div R_2(Y)$	The resulting relation R(X) includes all tuples t[X] in $R_1(Z)$ that appear in R_1 , in combination with every tuple from $R_2(Y)$, where Z= $X \cup Y$.

Consider for example the relations Painter and SketchArtist depicted in tablular form (Figure 3.8). The Painter relation contains a list of artists (identifiers) whose paintings are held in collections at the gallery and the SketchArtist relation contains a list of artists associated with sketches held in the archives at the gallery. Disregarding in this instance, all other artefacts (sculptures, photographs, manuscripts and other archival materials) held by the gallery, the curator requires the following information:

a) all artists for which the gallery holds paintings and/or sketches

b) artists for which the gallery holds *both* paintings and sketches

c) artists for which the gallery holds paintings only (no sketches are held for that artist)

d) artists for which the gallery holds sketches only (no paintings are held for the artist)

Painter	SketchArtist	a) UNION Painter ∪ SketchArtist	b) INTERSECTION Painter ∩ SketchArtist	c) DIFFERENCE Painter –SketchArtist	d) DIFFERENCE SketchArtist - Painter
ArtistID	**sArtistID**	**ArtistID**	**ArtistID**	**ArtistID**	**SArtistID**
10	**10**	10	10	20	42
20	**30**	20	30	50	140
30	**40**	30	40	60	160
40	**42**	40		70	
50	**140**	50			
60	**160**	60			
70		70			
		42			
		140			
		160			

Figure 3.8: Set theory operations: Union, Intersection & Difference on Painter and SketchArtist

Extending the UNION operation to an OUTER UNION operation allows partial type compatibility, where only some of the attributes in the tuples, match. Consequently, the resulting relation includes all tuples from both relations, including unmatched attributes and attributes contain either a value or NULL depending on the tuple match status (Figure 3.8a).

The CARTESIAN PRODUCT of the Painter and SketchArtist relations yields a relation comprising forty two tuples (Seven tuples in the Painter relation and 6 tuples in the SketchArtist relation) of degree two (Figure 3.9). Typically, on execution of this operation, subsequent operations to filter the resulting relation may yield more meaningful information.

Tuple / Row No.	**ArtistID**	**sArtistID**
1	10	**10**
2	10	**30**
3	10	**40**
4	10	**42**
5	10	**140**
6	10	**160**
7	20	**10**
8	20	**30**
9	20	**40**
10	20	**42**
11	20	**140**
12	*20*	*160*
13	*30*	*10*
14	*30*	*30*

Tuple / Row No.	**ArtistI D**	**sArtistID**
15	30	**40**
16	30	**42**
17	30	**140**
18	30	**160**
19	40	**10**
20	40	**30**
21	40	**40**
22	40	**42**
23	40	**140**
24	40	**160**
25	50	**10**
26	*50*	*30*
27	*50*	*40*
28	*50*	*42*

Tuple / Row No.	**ArtistID**	**sArtistID**
29	50	**140**
30	50	**160**
31	60	**10**
32	60	**30**
33	60	**40**
34	60	**42**
35	60	**140**
36	60	**160**
37	70	**10**
38	70	**30**
39	70	**40**
40	*70*	*42*
41	*70*	*140*
42	*70*	*160*

Figure 3.9: Cartesian Product: Painter X Sketcher, resulting in a new relation of 48 tuples.

Finally, consider the situation of a curator selecting paintings for an upcoming exhibition at the gallery and he needs to know whether or not particular paintings have previously been exhibited together, in the gallery. Given the relation ExhibitionDetail, containing all exhibitions and exhibits (identifiers) and the relation ArtworkSet, containing identifiers for each of the particular paintings of interest to the curator (Figure 3.10), the DIVISION operation answers this question. The result relation ExhibResult includes only those exhibitions (identifiers) in which all of the artwork of interest (ArtworkSet), were exhibited.

ExhibitionDetail

ArtworkID	ExhibitID
5	99
10	100
10	101
12	101
14	101
16	101
20	99
20	100
30	99
30	100
40	99
40	100
50	99
50	100
60	99
60	100
70	99
70	100
70	101
75	99
75	101
80	100

ArtworkSet

ArtworkID
10
20
70
80

EXHIBRESULT

ExhibitID
100

Figure 3.10 Division operation: Proposed paintings exhibited together once only, in ExhibitionID 100

3.3.1.2. Extending Set theory operations for the Relational Model

Extensions to Set theory algebra, required for the relational model, comprise SELECT (RESTRICT), PROJECT and JOIN operations (Table 3.2). Unary operations, SELECT and PROJECT enable data retrieval from one relation only. The SELECT (RESTRICT) operation Figure 3.11 SELECT (RESTRICT) unary operation defined for the relational model (Codd, 1970) returns a horizontal partition of the original relation, containing only all tuples satisfying specified conditions, in contrast to the PROJECT operation which returns a vertical partition of the original relation, containing only the specified list of attributes, for all tuples in the relation except discarded duplicate rows. The JOIN binary operations enable retrieval from two or more relations. JOIN operates on previously established relationships between entities (Foreign key – Primary key match). Consequently, NULL join attributes are treated as False and excluded from the resulting relation.

a) Consider, for example, that thousands of artists may be represented in the museum's collections, but that a curator is interested in only male, Irish artists. Providing appropriate selection conditions, logical operators and the original relation name in the SELECT operation (Figure 3.11), results in a relation comprising only tuples satisfying all conditions specified, that is, Artists that are male and that are Irish.

Executed operation: σ (Gender="M" AND Culture = "IRL") (Artist)

SELECT (also known as **RESTRICT:** is not the same as the SQL Select clause)

General Form: σ (selection condition) | operator | (selection condition) R

σ denotes SELECT

selection condition may be <attribute name> <comparison op> <constant value> or
<attribute name> <comparison op> <attribute value>

R is the name of specified relation

<comparison operator>: <, ≤, >, ≥, =,≠ and operators: AND, OR or NOT to combine selection conditions

SELECT operation is commutative: the conditions may be applied in any order
The degree (or number of attributes) of the resulting relation = the degree of R

Result relation contains tuples of all Male Irish artists and includes all attributes from the original relation

The **SQL** equivalent is SELECT * FROM ARTIST WHERE Gender="M" AND Culture = "IRL";

Figure 3.11 SELECT (RESTRICT) unary operation defined for the relational model (Codd, 1970)

b) Consider also the case where a curator wishes to print a list of the titles the year created, for all paintings in the collections. Specifying the required attribute names and the appropriate relation name, in the PROJECT statement (Figure 3.12), results in a relation containing the specified attribute names, from all tuples in the original relation, discarding duplicate tuples in the result set.

Executed operation: π (Title, Year) (Painting)

PROJECT

General Form: π <attribute list> R

π denotes PROJECT

<attribute list> contains the required attribute/column names in relation R, separated by a comma>

R is the name of specified relation

Result relation contains only the specified attributes, in the same order as in the attribute list, from all tuples in the artist relation, discarding duplicate rows.

The **SQL** equivalent is SELECT DISTINCT title, year FROM PAINTING;

Figure 3.12 PROJECT unary operation defined for the relational model (Codd, 1970)

c) Finally, consider the case where a curator requires a list of all paintings, showing the title and the year created, but he also wishes to include the artist name, birth date and gender. A JOIN operation may first be executed to combine all attributes of paintings and their associated artists, based on the established relationship '*Artist created a painting'* (Figure 2). However, since the curator requires a limited set of attributes, a PROJECT operation on the resulting relation from the JOIN operation, completes the report for the curator.

Format of the **JOIN** Operation: R1 ⋈ <join condition> R2

Executed operation: T1 ← PAINTING ⋈ <pCreatorID = CreatorID> ARTIST

Execute Project operaton $\pi_{\text{(Title, Year, Forename, Surname, BirthDate, Gender)}}$ (T1)

JOIN concatenates corresponding tuples from each relation into one longer tuple in the resulting relation. Therefore, if R1 has m attributes and R2 has n attributes the resulting relation will have a degree of m+n (attributes). Variations on the JOIN operation include a) *Theta Join* combines tuples from both relations, that satisfy the join condition, b) *EquiJoin* combines tuples from both relations that satisfy the join condition, but the join condition is restricted to the equality operator and c) *Natural Join* is the same as Equijoin except that the join attributes of R2 are excluded (the matching key values are already picked up from R1) and if the join attributes have the same name they do not have to be specified.

As described above, JOIN operations eliminate tuples that do not satisfy the Join condition, and are thus classified as type INNER JOIN. However, consider the case where a painting exists with no value set in the artistID attribute, but, the curator requires a list of all paintings, regardless of missing data. A Join, classified as an OUTER JOIN operations provide the solution, ensuring inclusion of tuples not matching the join condition, in the resulting relation. Executing a JOIN on relations Painting and Artist results in a relation containing three tuples, whereas executing a LEFT OUTER JOIN on the same relations includes non-matching tuples from the Painting relation (the leftmost relation specified in the operation), resulting in a relation containing five tuples (Figure 3.13). Similarly, executing a RIGHT OUTER JOIN includes non-matching tuples from the Artist relation (the rightmost relation specified in the operation). A FULL OUTER JOIN includes all tuples from all relations regardless of the join condition.

Necessary extensions to the original relational algebra, to satisfy reporting requirements in applications, allow additional operations on the PROJECT operation. By way of example, $\pi_{\text{<attribute list>}}$ R may be extended to $\pi_{\text{F1, F2,....Fn}}$ (R) where F1…Fn are functions. Given F1 as a function that calculates Freight costs increased by ten percent, the operation may be written as:

REPORT ← $^{\pi}$ FreighCost * 1.1 (STDCOSTS).

Additionally, tuples may be grouped by attribute values and aggregate functions, including SUM, COUNT, MINIMUM, MAXIMUM, AVERAGE may be applied to summarise data.

Artist

CreatorID	Surname	Forename	BirthDate	DeathDate	Gender	Culture	Etc	
JBY001	Yeats	Jack	29/08/1871	28/03/1957	M	IRL		
VER001	Vermeer	Johannas	01/10/1632	15/12/1675	M	NLD		
MAL001	Malevich	Kazimir	23/02/1878	15/05/1935	M			
KAH001	Kahlo	Frida	06/07/1907	13/07/1954	F			

Painting

ArtworkID	Title	Year	Technique	Medium	pCreatorID
001958	Men of Destiny	1946	OIL	CANV	JBY001
015001	Lady Writing a Letter with her Maid	1671	OIL	CANV	VER001
015985	The Geographer	1669	OIL	CANV	VER001
023456	Black Square	1915	OIL	LIN	NULL
023999	The Two Fridas	1939	OIL	CANV	NULL

JOIN (INNER JOIN)

$T1 \leftarrow PAINTING \bowtie_{<pCreatorID\ =\ CreatorID>} ARTIST$

PROJECT: $\pi_{(Title,\ Year,\ Surname,\ Gender)}\ (T1)$

Title	Year	Surname	Gender
Men of Destiny	1946	Yeats	M
Lady Writing a Letter with her Maid	1671	Vermeer	M
The Geographer	1669	Vermeer	M

LEFT OUTER JOIN

$T1 \leftarrow PAINTING \bowtie_{<pCreatorID\ =\ CreatorID>} ARTIST$

PROJECT: $\pi_{(Title,\ Year,\ Surname,\ Gender)}\ (T1)$

Title	Year	Surname	Gender
Men of Destiny	1946	Yeats	M
Lady Writing a Letter with her Maid	1671	Vermeer	M
The Geographer	1669	Vermeer	M
Black Square	1915	NULL	NULL
The Two Fridas	1939	NULL	NULL

Figure 3.13 INNER JOIN, LEFT OUTER JOIN, LOJ includes all tuples from the left relation (Painting)

Consider the case where a curator wishes to report the number of painting in each collection together with the temporal range of each collection, held in the museum (Figure 3.14). Combining both group and aggregate functions, the attribute list on the left hand side of the statement shows the reported attributes (renaming the original attributes) and the right hand shows the grouping attribute (CollTitle) followed by each of the selected attributes preceded by the aggregate function applied to it. **Statement:** $R_{(CollectionName,\ NoOfArtworks,\ EarliestYear,\ LatestYear)}\ ({}_{CollTitle}\ \mathcal{F}_{COUNT\ ArtworkID,\ MINIMUM\ Year,\ MAXIMUM\ Year}\ (PAINTING))$

CollectionName	NoOfArtworks	EarliestYear	LatestYear
Dutch	215	1560	1830
Irish	400	1800	1957

Figure 3.14: Extending relational algebra to enable Grouping and Aggregate functions.

Table 3.2 Relational Algebra operations developed specifically for the relational model.

Adapted from Codd, 1979 & Elmasri & Navathe 2014).

Operation	Type	Notation	Produces a Relation that:
SELECT (RESTRICT)	Unary	$\sigma_{\text{<selection condition>}}\ R$	Includes all rows matching the selection condition. All attributes in R are included.
PROJECT	Unary	$\pi_{\text{<attribute list>}}\ R$	Includes only those attributes from R that are specified in the attribute list, and in the same order as the attribute list.
(THETA) JOIN	Binary	$R_1 \bowtie_{\text{<join condition>}} R_2$ Comparison operators: $<, \leq, >, \geq, =, \neq$	Includes concatenations of all tuples from R_1 and R_2 that satisfy the join condition. The degree of the resulting relation = degree R_1 + degree R_2
EQUIJOIN	Binary	$R_1 \bowtie_{\text{<join condition>}} R_2$ OR $R_1 \bowtie_{\text{(<join attributes 1>), (<join attributes 2>)}} R_2$	Includes all combinations of tuples from R_1 and R_2 that satisfy the join condition, using *Equality* comparison only.
NATURAL JOIN	Binary	R_1 * <join condition> R_2 OR R_1 * (<join attributes 1>) , (<join attributes 2>) R_2 or R_1 * R_2	Same as Equijoin except that the join attributes of R_2 are excluded (the matching key values are already picked up from R_1). Join attributes with the same name do not have to be specified.

3.3.1.3. Relational Calculus

Codd (1971) proposed a second language or a 'sublanguage', an applied predicate calculus with tuple variables, to manipulate relations and to support storage and retrieval of formatted data from large shared databases. Described as a high level calculus oriented language, it refers only to names of: domains; database relations; workspace names; and tuple variables, in conjunction with a limited set of operations (Ibid.). Declarative statements include terms such as Get, Put, Insert, Update, Delete, Drop, query with qualification, declare a relation and its domains used in conjunction with the referred names surely lays the foundation of a higher level language such as Structure Query Language (Ibid.;Elmasri & Navathe, 2014, pp. 177-188). Queries in relational calculus rest on the concept of a tuple variable. A tuple variable ranges over a relation, such that the tuple variable stores the attributes and values of any tuple of the specified relation.

Consider the case where a curator requires a list of artist names and date of birth, for all Irish male artists in the system (Figure 3.15). Tuple variable *t* ranges over the relation ARTIST, therefore, all attributes prefixed by *t* refer to attributes from the ARTIST relation. The required attributes appear to the left of the expression, followed by a colon : and the condition or qualification expression comprising attribute names and the condition to be satisfied.

```
RANGE ARTIST t

GET W (t.forename, t.surname, t.birthdate):(t.Gender=”M” AND t.culture=”IRL”)

(W denotes a workspace on the user terminal or computer (Codd, 1971)
```

Figure 3.15 Relational calculus statement to list artist data. (Formatted according to Codd, 1971)

When more than one relation is involved in a query all tuple variables involved in the qualification or condition part and not in the leftmost (target list) must be quantified either explicitly or by default. The existential quantifier ∃ denotes *some* tuple variable while the universal quantifier ∀ denotes *all* tuple variables. By way of example, suppose a curator requires a list of painting titles and years, held in the collections for the artist (surname) “Yeats” (Figure 3.16). This statement is True if at least one or some paintings for the query Tuple variable *p* ranges over the Painting relation and tuple variable *a* ranges over the Artist relation.

```
GET W (p.title, p.year): ∃a(a.artistID = p.pArtistID AND a.surname = “Yeats” AND)

or

{ p.title, p.year | PAINTING(p) AND (∃a)(ARTIST(a) AND a.surname = “Yeats” AND a.artistID =
p.pArtistID }
```

Figure 3.16 Example GET / select query with bound Existenial quantifier a.surname

3.4. Relational Database Design Principles

Relational database (RDB) design principles promote sound methods towards achieving a robust and accurate machine representation and are particularly relevant to this study. The main concepts of the relational data model (Codd, 1970, 1971, 1979) distinguish it from previous models, such as the hierarchical and network models, particularly the insignificance of the order and position of data, in comparison to those models. Fundamental concepts to the relational database model comprise:

a) Domain: A domain is a set of values of a similar *type,* such as the domain of integers, the domain of part codes, the domain of characters, and it restricts the values allowed for the domain. In practice, a unique attribute name is assigned to each domain within a relation, depending on its role in the relation. By way of example, integers may be assigned to each ArtistID attribute ; a valid value in a BirthDate attribute may be a date in dd/mm/yyyy format.

b) Relation: in the mathematical sense, given sets S_1, S_2,…S_n (not necessarily distinct), R is a relation on these *n* sets, if it is a set of *n*-tuples each of which has its first element from S_1, its second from S_2 and so on to S_n, and where R is said to have *degree n*. In practice a relation represents only one concept or entity set of some particular aspect of the universe of discourse, for example, artist, artwork, style/movement.

c) Tuple: A tuple is a set of pairs (A:v), where A is an attribute and *v* is a value from the domain of A and represents an instance of an entity, represented by the relation. For example, ArtistID:10, ArtistSurname:Yeats, BirthDate: 29/08/1871, represents a particular artist in the Artist relation.

d) Primary Key: The function of a primary key (PK) is to uniquely identify each instance (tuple). A PK can be a simple domain (atomic attribute) or a composite key (two or more attributes). If more than one primary key is identified, that is, two or more candidate keys may uniquely identify an instance, then one primary key must be arbitrarily selected. Allied to the concepts of primary key is the concept of a superkey, where a superkey (a combination of one or more attributes) uniquely identifies an instance. While all tuples have at least one default superkey, (comprising all attributes), one or more superkeys may be present. A candidate key on the other hand, represents a minimal superkey on the relation. Since relationships (especially other than IS-A relationships) between entities require a foreign reference, composite keys may degrade database performance because all fields involved in the PK must be copied to other tuples (as a Foreign key). Furthermore, the RDBMS must check all key fields when data partaking in the relationship changes through insert, update or delete operations. Where large composite keys are required, a Surrogate key may provide a better solution, especially where the database automatically generates a unique number for each instance added to the relation.

e) Foreign Key: A foreign key establishes a relationship between relations (or tuples within a relation) by cross reference to a primary key in another tuple. Therefore, a foreign key is an attribute whose value is the same as the value of a Primary Key in another relation. Similar to primary keys, foreign keys can be single or multi attribute keys.

f) Views: A view provides a means of data abstraction by enabling aggregation, combination of relations or selection of subsets of data, derived from the underlying base tables, into a virtual table, that remains empty until materialized or populated in response to a query. By way of example, if 'genre' (denoting, landscape, historical art, still life, portrait etc) is an attribute in the artwork relation then a named view may be created to refer to a particular genre. Therefore vwLandscape would comprise only landscape paintings, vwPortrait would comprise only Portrait paintings and so forth. Furthermore, each view may combine attributes from different relations, for example, both the Artwork relation and the Artist relation which enables, for example, automatic retrieval of the Artist name, birth/death years and nationality along with the artwork details, without a programmer having to specify these details.

Regarding relationships between entities, consider, for example, a relationship between an artist and a painting and the complex relationship that is an otherwise intuitive description, illustrated in tabular form (Figure 3.17). The Artist relation and the Painting relation describe two different entity types, central to the cultural heritage (CH) domain. Since the artist relation specifically describes artists, the relation comprises attributes of interest to the domain, about an artist only, while a different set of attributes describe each artwork in the Painting relation. Establishing the relationship '*Artist created a Painting'*, requires a foreign key(pCreatorID) in the Painting entity, that references the primary key(CreatorID) in the Artist entity.

Artist

CreatorID	**Surname**	**Forename**	**BirthDate**	**DeathDate**	**Gender**	**Culture**	**etc**	
JBY001	Yeats	Jack	29/08/1871	28/03/1957	M	IRL		
VER001	Vermeer	Johannas	01/10/1632	15/12/1675	M	NLD		
DAL001	Picasso	Pablo	25/10/1881	08/04/1973	M	ESP		

Painting

ArtworkID	**Title**	**Year**	**Technique**	**Medium**	**PCreatorID**
001958	Men of Destiny	1946	OIL	CANV	JBY001
015001	Lady Writing a Letter with her Maid	1671	OIL	CANV	VER001
015985	The Geographer	1669	OIL	CANV	VER001

Figure 3.17: Relationship between Artist and a Painting.

CreatorID, the Primary key on the Artist table, copied to Painting table where it is a Foreign Key back to the Artist table.

Adopting common terminology, metadata standards and annotation standards developed for the CH domain (Section 2.3) promotes a design reflecting the shared semantics of the domain and contributes to rapid understanding and usability of the model, in addition to supporting future interoperability. For example, object name appears in AAT, Getty vocabulary as ID: 300033618, Painting (visual work); similarly, Painters appears in AAT as ID: 300025136, Painters (Artists). Attribute values may also contain common values defined in CH and or international standards across all domains, as illustrated in the foreign key attribute 'Culture' in the Artist table, where ISO 3166 is the International Standard for country codes and codes for their subdivisions [37]. ISO 8601 Time intervals and time points as implemented by schema.org

[37] https://www.iso.org/iso-3166-country-codes.html Last viewed 18/09/2017

3.4.1. Database integrity and constraints

Database integrity is enabled through constraints defined on the database. A number of constraints or rules are available to the database designer to maintain database consistency (Coronel & Morris, 2016, pp. 78-81; Elmasri & Navathe, 2014, pp. 69-90, 131-133). Constraints defined during data definition include: A) the NOT NULL constraint: Since a primary key uniquely identifies an entity the value of a primary key cannot be NULL. This *entity integrity* constraint means that each tuple in a relation can be distinguished and referred to by other relations and by SQL operations.

b) Referential integrity constraints operate between two relations. Referential Integrity means that a foreign key cannot contain a value that does not exist as a primary key in a corresponding table. However, the value of a foreign key may be NULL until such time as it is allocated to a primary key.

c) *Attribute constraints* control the values that can be stored for an attribute. d) The *NOT NULL constraint* ensures that insert and update operations will be successful only when a value is allocated to attributes defined as NOT NULL. e) The *Default* constraint automatically stores a default value to an attribute when a value is not otherwise specified by the user or an operation on the database. f) A *Check* constraint allows only those values specified in a list of valid values to be stored in the database. Alternatively, a *check* constraint may define a rule limiting the value allowed, for example, check (DeathDate >= BirthDate OR NULL). While a primary key automatically guarantees uniqueness of an attribute (or set of attributes if a composite key) a g) *Unique* constraint may be defined on attributes to ensure that any attributes other than the primary key are unique.

In addition to the constraints defined in data definition language when tables are created and altered, *business constraints* may be applied to the database using Assertions and Triggers. A trigger is defined on a relation and comprises procedural code that fires automatically, when some event (such as Insert or update), occurs. Assertions are statements defined on a database that check adherence to the application or business rule.

3.4.2. Functional Dependencies and Normalisation

The concept of functional dependency (FD) is fundamental to RDB design theory, and relies upon the designer's understanding of the semantics of the relations and their attributes (Elmasri & Navathe, 2014, pp. 520-523). FD denotes a constraint between attributes in a relation identifying attributes determined by one or more other attributes in the relation. By way of example, if a museum allocates a unique number to each painting in the collection, then that unique number or 'primary key' identifies each individual painting and determines all other attributes of the painting. Moreover, if a curator knows the unique identifier of a painting, he can discover all characteristics of the particular painting based on a search by unique identifier. In reality, multiple FD's may exist on a relation (Figure 3.18). All attributes, except ArtistName are determined by artworkID in FD1. Alternatively, FD2 presents all attributes determined by the title of the artwork. However, given that there is no guarantee of the uniqueness of titles or names, FD1 represents a more viable option. The third FD, depicting aArtistID as the determinant of ArtistName highlights an incorrect structure because it is an attribute of the Artist and does not describe the artwork. As such, violates Normal Forms, 2NF (Table 3.3)

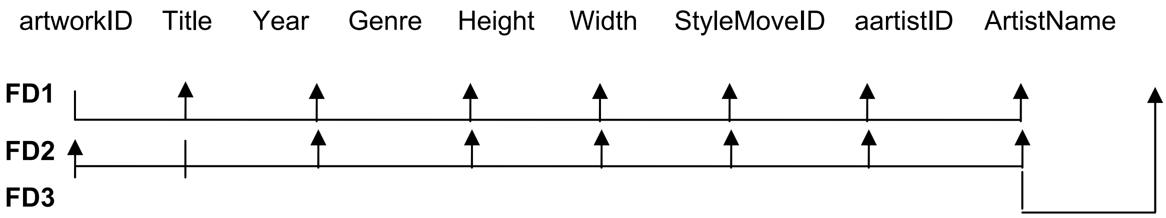


Figure 3.18 Functional Dependencies on the Painting relation.

Perpendicular line with no arrowhead denotes the determinant or identifier, lines with arrowheads denote the dependant attributes.

Robust relational design involves *Normalization*, an iterative process to design correctly structured relations (Codd, 1970), for persistent representation. Correctly structured relations comprise distinct tuples, determined by a unique identifier, representing one instance of an entity (type); comprising atomic attributes and where row order and column (attribute) order are insignificant.

Table 3.3 Normal Forms, rules for designing correct structures for the relational model

Normal Form	**Requirements**
1NF	Each tuple is unique and all tuples have the same single-valued, atomic attributes.
2NF	Relation is in 1NF and every nonprime attribute is fully functionally dependent on the primary key (PK), that is, on all parts of the PK if the PK is a composite key.
3NF	Relation is in 2NF and no nonprime attribute is transitively dependant on the PK, that is, no nonprime attribute is a fact about another nonprime attribute
BCNF (Boyce Codd)	Relation is in 3NF and all determinants are keys, that is, only *one* possible primary key exists in the relation.
4NF	Relation is in 3NF and the relation does not contain two or more *independent* multi-valued facts about the key Example: If each artwork may be associated with many other artworks and each artwork may also be associated with many keywords then these facts should be defined in two new relations: R_1 artwork-artwork and R_2 artwork-keyword
5NF/PJNF (PROJECT-JOIN)	Relation is in 4NF and join dependencies are based on candidate keys. Same as 4NF unless some symmetric constraint exists. If constraint exists then original relation cannot be recreated from the decomposed relations.
6NF	A relation in 5NF + no non-trivial join dependencies, including temporal data.
DKNF (Domain Key)	A relation in which all constraints are functions of domains and primary keys

Furthermore, benefits of normalization include a schema that accurately describes the universe of discourse and lays the foundation for data consistency, minimized redundancy and future development that precludes insert, delete and update anomalies. Normalization proceeds by analyzing a given schema or an initial intuitive set of entities and applying increasingly strict tests (Table 3.3) to identify candidate keys, for selection of the primary keys and establishing a set of base relations and relationships to other entities (Fagin, 1977, 1981; Kent, 1983).

Functional dependencies and normalization are applied in the next section that describes the process of designing a database for our study.

3.5. Relational Database Design

This section describes the steps involved in designing a relational database. For this research, the universe of discourse is the cultural heritage domain, but similar database design procedures and techniques apply, regardless of the specific domain except that the designer adopts the relevant metadata and content standards for that domain, if they exist, or otherwise develops new metadata schema as appropriate.

The database design process consists of three components: a) logical modeling and c) application of domain content standards and c) database schema creation

- The logical modeling component involves a mapping process from a conceptual model to a logical model, also represented in graphical form, in an Entity Relationship Diagram (ERD). The input to this process includes the previously designed conceptual model (Section 3.2)

- Application of domain content standards: Since the content of many elements across the domain employ standard formats, ISO standards and/or vocabularies, deferring discussion of these matters is more appropriate at the end of the logical modeling component and avoids repetition.

- Creating the database schema involves designing SQL DDL scripts (Structured Query Language, Data Definition Language) and running these scripts to build the database schema, based on the logical model output from the mapping process

3.5.1. Logical Modeling – Creating a database

Creating a logical model of the database builds on the methodology employed in conceptual modeling (3.2) and is fundamental to successful design of the database schema for the domain. The logical modeling process, examines the abstract constructs represented in the conceptual model – the *entity types*, *relationships* between the entity types, the *attributes* (attributes being minimally defined in this case) and the *generalization/specialisation*, and leads to further refinement in order to create an accurate representation of the domain, at a more detailed level. The process again involves Entity relationship modeling and established and proven mapping procedures and techniques developed by database experts (Chen, 1976; Elmasri & Navathe, 2014, pp. 287-301; Teorey et al., 1986) to map constructs from the conceptual model to a logical model. The mapping process necessarily includes functional dependency (FD) analysis and normalization techniques and results in splitting relations into several smaller relations, ensuring the database schema contains well defined structures. Adhering to established database design principles also reduces potential sources of insert, delete and update anomalies.

The mapping process desBased on the requirements for this study, inputs to the mapping process comprise:

- The conceptual model (3.2)
- Findings from the mixed methods study (Chapter 4.)
- Metadata Standards: Categories for the Description of Artworks (CDWA) and the Visual Resources Association (VRA) Core 4.0 (Baca & Harpring, 2014 -Revised online 2016; VRA, 2007)
- A content standard: Cataloging Cultural Objects-CCO (Baca et al., 2006)

The traditional mapping procedure comprises the following steps:

1 Map regular entity types
2 Map weak entity types
3 Map binary 1:1 (one-to-one) relationship types
4 Map binary 1:N (one-to-many) relationship types
5 Map N:M (many-to-many) relationship types
6 Map multivalued attributes
7 Map N-ary (N>2) relationship types
8 Map specialization or generalization

Given the complex nature of the entity types in the cultural heritage domain, for this exercise the designer first examined the core entities and relationships and completed all steps of the procedure for each one, as appropriate, before proceeding to the next core entity type. The following sections describe mapping the core entity types for the domain.

3.5.1.1. Map Regular entity types – Agent

Core entity types of interest in the domain comprise Agent, Work, and Archives. The Agent entity type contains records for individual persons, artists, and corporate bodies, including recognized artist studios. Similarly, the Work entity type contains individual records for each work of art. On the other hand, a collection such as archives, representing a number of objects from one source, or shared provenance, and organized in some logical structure, is the subject of the next section (3.6) describing the archives harvesting process.

Starting with the Agent entity type, corresponding to a Personal and Corporate Name Authority (Baca et al., 2006, pp. 279-311) and incorporating ULAN vocabulary, this entity type comprises the museums' records of individuals and corporate bodies associated with the arts. The Agent entity type holds details of artists and all non-fictional persons and organizations related to the artist, such as, noted family members, patrons, mentors and other acquaintances. It may also store details of domain specialists such as prominent academics and art historians along with institutions, museums, repositories, libraries and other organizations associated with the arts and archival materials.

Examination of the items describing the Agent/Artist (Table 3.4) informed by the metadata and cataloging standards, from a database design perspective, reveals that storing the data as depicted, would not meet normalization standards.

Table 3.4 The Agent entity

Agent	**Example**
AgentID	Unique number allocated by the DB or a code allocated by the cataloguer
Type (Person or Corporate body)	Artist
Name	Jack B Yeats, W Bird (pseudonym– as illustrator, graphic artist)
Biography	Text description
Birth Date	29 August 1871, London, England
Death Date	28 March 1957 (aged 85), Dublin, Ireland
Nationality/Culture/Race	Irish
Life Roles	Artist, Painter, Illustrator / graphic artist, Author
Name Source	Refer to a bibliographic source
Date of Earliest Activity	1888
Death of latest Activity	1956
Place/Location	Place associated with particular events?
Gender (optional)	Male
Related People /Corporate Bodies	John Yeats,-Artist, W B Yeats Poet, other family members, John Millington Synge, Samuel Beckett, Oskar Kokoschka
Relationship Type	Father, Brother, Mentor, Influenced by, Influence on etc.
Events	Yearly visits to Ireland, collaboration with J.M Synge, Authored books and plays. Elected member of RHA in 1916
Note	Any free form text describing the Agent, not covered elsewhere

Closer examination of the Agent entity type also takes into account, recommendations from the content standards (Baca et al., 2006, pp. 78, 282-283) that state:

- "*The creator's names, nationality, life roles, and life dates are critical access points*"
- "*Display Biography, Birth Date, Death Date, Note, and Gender should not be repeatable elements. All other elements should be repeatable.*"

Reviewing the listed elements (Table 3.4) reveals evidence of repeating groups of data or multi-valued attributes, and more non-atomic values, compelling several adjustments to achieve 1NF (first normal form). By way of example, *Name*, *Life roles, related people* and *events* all contain repeating elements and *BirthDate and DeathDate* represent composite elements, containing different types of values: dates and locations. 1NF rules require removal of repeating elements, and reducing composite elements to atomic elements.

However, in the cultural heritage (CH) domain artist names are recorded in one attribute, by way of example, agentName may be stored in natural order “Jack B Yeats”, (display name) whereas the index to support search and retrieval is generally stored in inverted format “Yeats, Jack B” (Baca et al., 2006, p. 289). Alternatively the supporting index may contain the commonly accepted name for renowned artists, by way of example ‘Rembrandt’ and not “van Rijn, Rembrandt”. This is standard for the CH domain especially for easier since search and retrieval by art experts and the general public who use the commonly accepted name.

Also of note in the CH domain is the fact that dates are generally recorded as year only and do not include day and month. Given the temporal range of CH, exact dates may not be known and do not adversely affect the data, unlike for example medical systems or finance systems where exact dates may be critical to their operations.

On its own, *Place/Location* is ambiguous and without clarification, it cannot be fully functionally dependent on the Agent. The likelihood is that *place* refers to where some event occurred, rather than describing the Agent and at this point, removal from the Agent entity type is the most prudent action to take (2NF).

The *Relationship Type* entity does not describe a particular characteristic of an Agent. Rather, it exists outside of the Agent and defines the different types of relationships between the Agent and other Agents. Also, *Role* describes the Agents role or contribution to some activity as opposed to a characteristic of the Agent. By way of example, Jack B Yeats acted in the role of *artist* (painter) for the painting ‘For the Road’; in the role of *illustrator* for Boys Own Paper, in the role of freelance *writer* for Punch magazine and as *author* of several books and plays. Therefore, as multi-valued descriptors for entity types related to the Agent and not innate attributes of an Agent, removal of these elements from the Agent record satisfies 3NF rules.

Consideration of the *Nationality* item shows that the term 'Irish' describes the Agent’s nationality. However, duplication of this term when describing other Agents increases the probability of inconsistent spelling (update anomalies). Similarly, amending the term from ‘Irish’ to ‘Ireland’ requires amending every Agent record containing the term ‘Irish’, to avoid similar anomalies. Finally, removing the last Agent from the database, containing the term ‘Irish’ would result in losing the term ‘Irish’. That being said, the conceptual model already depicts Place as a regular entity, corresponding somewhat to the Geographic Place Authority (Baca et al., 2006, pp. 312-330).

Having removed the repeating groups, functional dependency (FD) analysis identifies a combination of Agent Name and Birth Date as a possible candidate key, in view of the fact that Agent Name on its own may not guarantee uniqueness across all tuples/records in an Agent relation (table). However, given that the Agent may be involved in many different relationships, as described above, it is more efficient to retain the one attribute primary key (PK) designated during the conceptual modeling phase rather than forcing each relationship with the agent, to store multiple attributes of a composite key, as a foreign key to the Agent.

Continuing deliberations on the Agent entity type, the concept of unknown artists requires attention. Arising from our interviews with art experts (Chapter 4) is the fact that an artist or creator of a work of art may be unknown. Suggested workarounds for this problem, rather than accept a work of art with no creator specified (using a NULL value for an artist), include designations that contain the location and or time span, such as *Florentine* or *unknown 16th-century Florentine*, (Baca et al., 2006, p. 280) or perhaps a combination of 'unknown' and culture, for example *unknown German*. However, using generic terms as suggested above means that such agents may create multiple different works of art and the authors recommend that museums devise a standard and consistent approach to address the issue of unknown artists.

Furthermore, the interviews revealed that there may be more than one artist/creator associated with a work of art and that each one may contribute in a different role. Therefore, rather than establishing a relationship directly between a work of art and the agent entity, an intermediary relation is needed, such as an 'Artist-role' entity.

At this stage, the process results in nine entity types, corresponding to nine relations (tables) in the database schema. These comprise: Agent, RoleType, Place, RelationshipType, and EventType representing regular entity types, whereas, entity types, AgentLnkName, AgentLnkRole, AgentLnkAgent and AgentLnkEvent represent multi-valued weak entities (Figure 3.19, Table 3.5). Derived entity types based on the repeating groups of Names, Roles, Relationships and Events represent *weak* entity types, whose existence depends on the presence of an owner or parent entity type. Moreover, the weak entity type always contains a foreign key (FK) equivalent to the parent primary key (PK) of the Agent, satisfying referential integrity constraints. By way of example, removing the one and only painting created by Gabreil Metsu, from the collection and removing the artist Gabriel Metsu from the Agent relation, compels removal of the roles, events, relationships and names associated with Gabriel Metsu.

A closer look at each weak entity is necessary before deciding on the PK, the degree of the relationship between dependents and owners, and if their relationships are mandatory or optional.

- An Agent has at least one name (or allocated a name as discussed above), but may have multiple names, this represents a mandatory 1 to N (many) relationship between the Agent and the AgentLnkName *weak* entity.
- Similarly, an Agent participates in at least one event, either the birth of the artist or creation of a corporate body, and therefore the Agent has a 1 to N relationship with the AgentLnkEvent weak entity.

- An Agent may or may not be associated with one or more Agents, therefore the relationship between the Agent and the AgentLnkAgent represents an optional 1 to N relationship.

- The role element describes the life role and/or the contribution to the creation of a work of art, regarding an artist Agent. For other individuals lacking a defined role, the system may allocate a default role. For a corporate body the role denotes the function of the Agent. Therefore, the relation between Agent and AgentRole represents a 1 to N relationship.

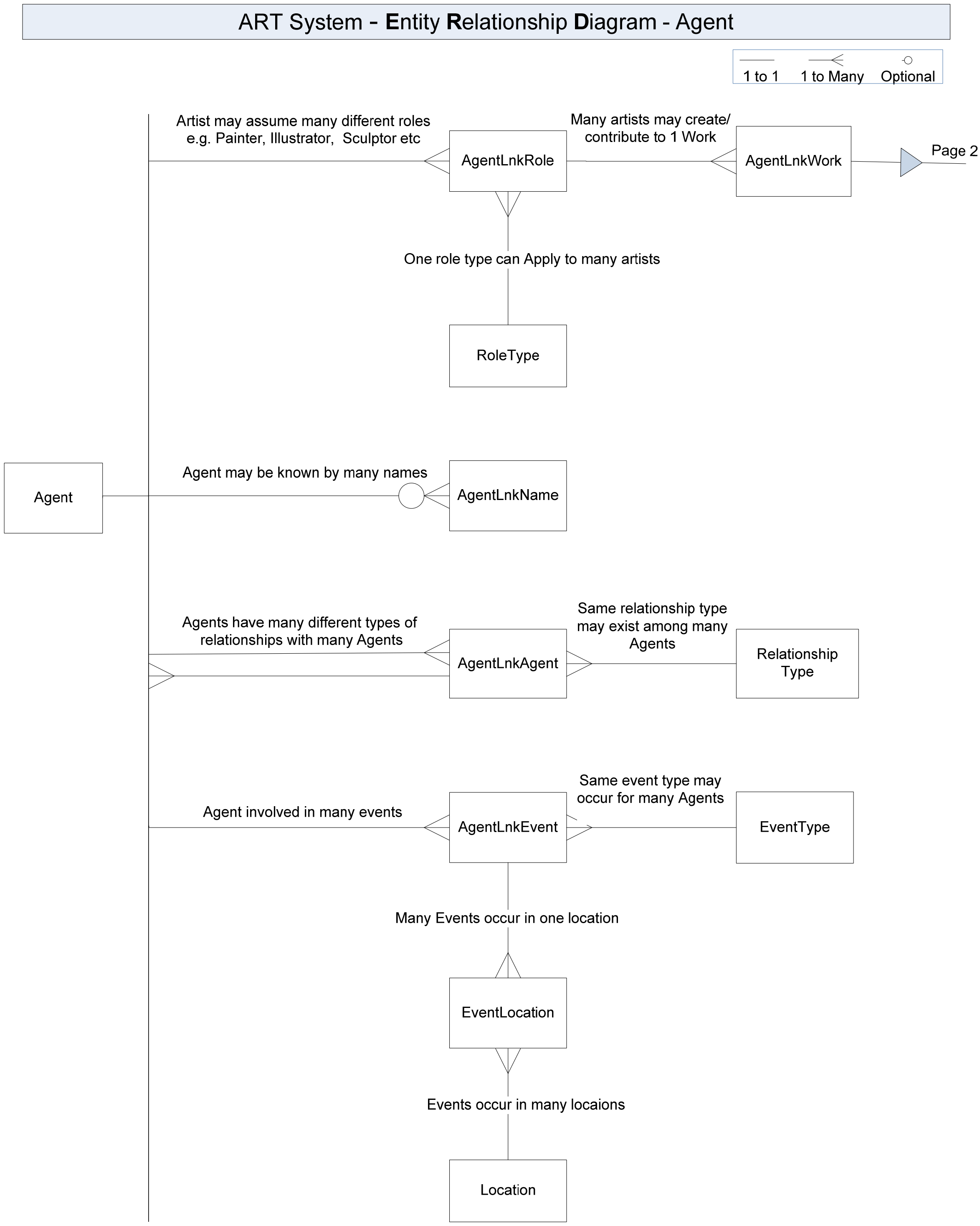


Figure 3.19 Entity Relationship Diagram (ERD) for Agent and associated relations.

Table 3.5 Database schema – result of logical modeling of Agent

Entity Type	**Primary Key**	**Attributes->**									
Agent	AgentID	DisplayName	BirthDate	DeathDate	EarliestDate	LatestDate	NationID	Gender	Biography	Note	Index-Name
AgentType	AgentTypeID	AgentTypeName	Vocab	Refid							
RoleType	RoleID	RoleName	Vocab	Refid							
PlaceLocation	PlaceID	PlaceName	Vocab	Refid							
RelationshipType	reltypeID	relTypeName	Vocab	Refid							
EventType	EventTypeID	EventName	Vocab	Refid							
* AgentLnkName	AgentLnkNameID	AgentID	AgentName	PrefferedYN	Lang						
* AgentLnkRole	AgentLnkRoleID	AgentID	RoleID	roleNote							
* AgentLnkAgent	AgentLnkAgentID	Agent1ID	Agent2ID	reltypeID	relNote						
* AgentLnkEvent	AgentLnkEventID	AgentID	EventTypeID	EventDate	EventNote						

3.6. Harvesting archives from texts

The aim of this study is to design a computational representation that incorporates the knowledge of domain experts such as curators. This representation is based not solely upon works of art, but upon the combination of works of art and associated collateral materials from other primary sources, such as artist archives. Containing contemporaneous unique materials created by the artist or accumulated over the course of a lifetimes work, domain specialist recognize artist's archives as potentially rich resources when researching a work of art (Breakell, 2015; M. Reed, 2017; Sjöholm, 2014). Millions of collectible objects from sketchbooks, to photographs, to diaries, to letters, to artist tools and miscellaneous ephemera, reside in artist archives. These materials are arranged into special collections (Fonds) and preserved and documented by museums.

This section describes the process and methods for harvesting archives from texts and storing these in a database, for future search and retrieval and underpinning applications for domain specialists when researching an artwork. The particular texts for this research include documentation of archival materials preserved in the National Gallery of Ireland. The archives are documented in PDF files, described in accordance with a standard for archives used in the cultural heritage domain: the General International Standards for Archival Description (ISAD(G)).

The structured nature of these texts affords the opportunity for a computational method of extracting details about the archives. Harvesting this structured data using a computational process allows large volumes of data to be harvested automatically. Automatic harvesting relieves the burden of manual data entry. Moreover, it retains the essential contextual information and the hierarchical structure of the archives, in compliance with archival principles. Furthermore, harvesting the archives into a persistent representation that also adheres to a metadata standard, promotes consistency of description, which is fundamental to future interoperability.

Academic research, tools and techniques in the text analysis arena, such as the Stanford Core NLP Toolkit (Manning et al., 2014), inform the process and methods for harvesting archives from texts. Despite the fact that this harvesting process collects data from texts organized accorded to ISADG, and that the texts relate solely to the artist Jack B Yeats, similar analysis and extraction principles apply when harvesting data from different texts and/or metadata standards. With some adaptation the harvest program may be reused. The overall approach adopted for this process employs the following tools and techniques:-

a) A glossary of terms (based on the metadata standard of the source texts)
b) Term recognition (based on a glossary of terms)
c) Term positioning to locate key terms and corresponding values

d) Term frequency (to validate / count the sections and materials captured)
e) Regular expressions for pattern matching and source data cleansing and reformatting
f) A recognized metadata standard for the destination persistent storage of archives

The harvesting process consists of two components: a) Textual analysis and b) Archives extraction into a persistent storage model.

- The textual analysis component takes text files of structured data as input and pre-processes the data by cleaning and reformatting content using pattern matching techniques. It also detects and stores occurrences and positions of glossary terms within the texts.

- The archives extraction component extracts content by calculating the distances between the terms in the texts and storing the intervening strings into metadata elements in the database.

Before describing the harvesting process in detail, examination of the content and structure of the source texts provides an indication of the types of data and structures involved.

3.6.1. Source texts

As previously noted, the source texts input to the harvesting process are organized according to archival principles, including the principles of *provenance* and *respect des fonds* (Duchein, 1992; Sweeney, 2008) that emerged with archival science in the seventeenth century. Despite differences of opinion and challenges to these principles (Cook, 1997) especially with respect to personal papers and usability of these materials by those accessing the archives (Boles, 1982), these principles are upheld by ISAD(G) and Encoded Archival Description (EAD) standards.

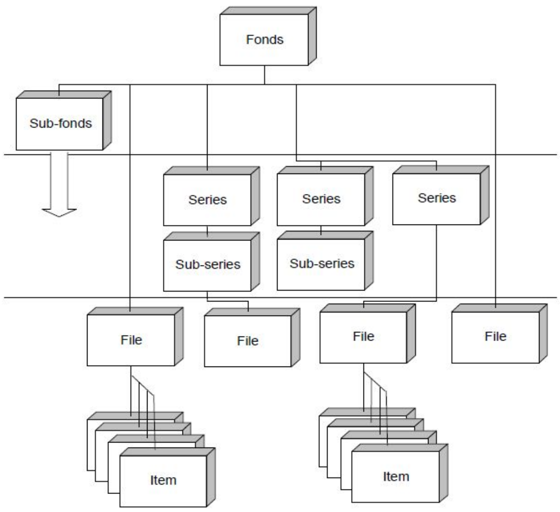


Figure 3.20 Model of the levels of arrangement of a fonds. (ICA,2000 p.36)

According to these principles all of the records in the collection (or more correctly, the *Fonds*) are from the same entity or source (provenance). Additionally, the records are arranged in the original order (respect des fonds) as organized by the creator. Arranged in a hierarchical structure (Figure 3.20) the *Fonds* or top level describes the whole of the archive, that is, all of the materials described in all sections of the document. A *series* describes materials held in some filing system grouped according to similar activity, form, or a particular relationship among the items. A *file* represents an organized group of artefacts used by the creator or grouped by similar subject or activity and an *item* represents the lowest level; an intellectually indivisible unit such as a sketchbook or a letter. ISAD(G) documentation states that all higher level descriptions should be accessible from all lower level records.

The International Standards for Archival Description (ISAD(G)) metadata and content standard identifies and defines seven categories of information, each comprising a set of selected details from a total of twenty six elements (Table 3.6) (ICA, 2000).

Table 3.6 ISAD(G) Categories and details to describe archival materials

Including metadata elements of the selected storage model: Encoded Archival Description (EAD)

Category / Area	Details	EAD Fieldname
1. Identity Statement	* Reference Code	eadid countrycode (IE)
	* Title	unittitle
	* Date(s)	unitdate
	* Level of Description	unitdesc
	* Extent & Medium of the unit of description (quantity, bulk or size)	Extent
2. Context	Name of creator(s)	origination
	*** Administrative/Biographical History	bioghist
	Archival history	custodhist
	Immediate source of acquisition or transfer.	acqinfo
3. Content & Structure	** Scope and Content	scopecontent
	Appraisal Destruction and scheduling information	appraisal
	Accruals	accruals
	*** System of arrangement	arrangement
4. Conditions of Access & Use	Conditions governing access	accessrestrict
	Conditions governing reproduction	userrestrict
	Language/scripts of materials	langmaterial
	Physical characteristics and Technical Requirements	phystech
	Finding aids	otherfindaid
5. Allied Materials	Existence and location of originals	originalsloc
	Existence and location of copies	altformavail
	Related units of description	relatedmaterial
	publication note	bibliography
6. Notes	Notes	Note
7. Description Control	*** Archivist's note	processinfo
	*** Rules or Conventions	descrules
	*** Date(s) of Descriptions	processdate

Content requirements and mandatory or optional entries, depending on the level of description, means that not all descriptions require the same level of detail. Therefore, all records do not have entries for all twenty six elements. Mandatory and optional elements for all levels of description are denoted: an asterisk * and double asterisk ** denotes a mandatory element at all levels of description as adapted to Irish standards and *** denotes mandatory elements required at the 'Fonds' level of description, as adapted to Irish standards. Authority details are contained in particular metadata elements also, including 'Name of creator(s)'; 'Administrative/Biographical History' and 'Archivists Note' at the 'Fonds' level of description. However, other names included might be those associated with the creating entity and / or particular items, for instance the author or recipient of a letter.

3.6.2. Document and database preparation

Having selected the source documents from which to harvest the archives, the first preparatory step required conversion of these PDF documents to text UTF-8 (Unicode Transformation Format) files; so that they could then be processed by the harvesting program. Using the Save As function from Adobe PDF allows the user to specify the document type (plain text) and the encoding format.

Next, the glossary of terms that would initiate the extraction process, needed to be acquired. Since there are only twenty six ISAD(G) terms to consider, the glossary was compiled using an SQL (Structured Query Language) script to populate a 'glossary' table, which was also created by an SQL script. An alternative method, especially for a large volume of terms, could have been to read in the terms from a file or a database. Two extra values are added to the glossary of terms: one determines whether finding this term forces a new record to be inserted into the database, and the other field denotes the destination fieldname into which the associated term value, will be stored. These two fields could have been placed in a separate table and a View based on the UNION of the glossary table and that table would have achieved the same end. That is, to pick up the new record flag and the field name. Other tables to store the harvested records and temporary processing tables were also created using SQL scripts run on the database.

The database used for this study is MySQL and all scripts were run directly on the database using the phpMyAdmin tool: both of which are free tools and available for download from the web.

Initial attempts at harvesting the archives revealed several anomalies. Differences in terminology between the glossary and the texts and among the texts at different levels of description, resulted in concatenation of values and incomplete records. For example, the term 'Administrative/Biographical History' appeared in a variety of forms in the texts ranging from 'Biographical note' to 'Biographical history' to 'Biographical history:' to 'Biographical', to 'Biographical/Administrative history'. Similar variations in terminology were found for several other terms and some included a colon as the last character, others included a space and colon while other terms did not include the colon. Additionally, the term ‘Extent and Medium:’ appears uniformly across the documents at the ‘Fonds’ level, in place of the standard term ‘Extent & Medium of the unit of description (quantity, bulk or size)’ but was omitted from other levels. Inspection of the texts also highlighted errors in the sequence or order of the descriptions. For example, ‘Series’ level of description appeared after the actual title of the series and its reference code was missing. However, these were implied by the subsequent items described below it. Data truncation also occurred as some values (Biographical history) were too large for the destination field definitions. To the human eye all the above anomalies (apart from data truncation) do not pose a problem and people implicitly understand their meaning, but a computer only detects strings as instructed by the program. Through trial and error the resulting solution comprised of:

a) Manual search and replace of terms in the input texts, adding :: to mark the end of the term

b) Manual change to move the series title after the term ‘Series’ and put in the reference code for the series (derived from the records within the series)

c) Altering the glossary table to provide up to four alternative formats of the term.

d) Adjusting the program to include Regular Expressions for pattern matching to automatically insert missing terms, as needed.

e) Use pattern matching to parse strings in order to extract specific data.

f) The *Title* field and *Scope and Content* field are sometimes used interchangeably. If there is no title then the program will take the first sentence only from the scope and content, as the title. All sentences in the section will be mapped to the scope and content field.

g) Dates are not readily available for indexing but are extracted as given. When designing the application indexing dates (for searching and sorting) will be generated by applying rules adopted by the National Gallery of Ireland.

3.6.3. Harvesting archives

Having built the required tables in the database and prepared the input texts, the next stage is the text harvesting pipeline. This commences with a pre-processing stage, the purpose of which is to transform the texts data into a format that can be stored in a database. To prepare the data for insertion to a database a number of processes are applied to the input texts:

a) Pattern matching
b) Term location calculation

Once the ‘Term Location Calculation’ step is completed the next step in the pipeline is to continue with harvesting values to store in the database. This involves two steps:

a) Extract key value pairs
b) Value normalization or cleansing

The four steps of the harvesting process are described below.

3.6.3.1. Pattern Matching

Initial attempts to harvest the archives revealed that some terms had been omitted from the documents. To rectify this problem a series of regular expressions were defined in the program. These regular expression or patterns (Table 3.7) are applied to the input texts and reformat the text by inserting the implied, but omitted, terms. By way of example, the text (Figure 3.21) represents some level of description; assumed to be ‘Item’ level of description but the terms Level of Description, Dates of Creation, Extant and Medium and Scope and Content are missing

Y2/1/1 12 April 1954 1 item, 2pp Letter from Jack Butler Yeats, Portobello house, Portobello bridge, Dublin, to Eleanor de Bretteville Reid. Yeats thanks Reid for gifts sent to him and for a ship sent to his nurse’s son which has not arrived yet. He is glad she liked [...]. Asks that Frederick W. Reid send him a copy of [his poem] ‘Stardreams and pigments’ for a friend.

Figure 3.21 Sample for the input texts. [...] denotes removal of text for purposes of this example

To the trained eye of the archivist, this (Figure 3.21) reads as: ‘Reference code: Y2/1/1’, it is an ‘Item’ level of description, the ‘Date(s) of Creation’ is 12 April 1954, and the ‘Extent and Medium’ is ‘1 item, 2pp’. Applying the regular expressions in examples 1 and example 2 (Table 3.7) mimics the expert eye and automatically places the required terms before their associated values, in the text file.

Table 3.7 Pattern matching examples to output texts to include Terms missing from the input texts
Pattern matching examples to output texts to include Terms missing from the input texts

Ex. 1 RegEx Pattern	**(".$CollectionID.")(/)(\d+)(/)(\d+)**
Pattern replacement	Reference Code:: $1$2$3$4$5 Dates of Creation::
Input text	Y2/1/2
Output text	Reference Code:: Y2/1/2 Dates of Creation::
Ex. 2 RegEx Pattern	**(\d+)(item,)(\d+)(p+)**
Pattern replacement	Level of Description:: Item Extent and Medium:: $1 $2 $3 $4 Scope and Content::
Input text	1 item, 2 pp
Output text	Level of Description:: Item Extent and Medium:: 1 item, 2pp Scope and Content::
Ex. 2 RegEx Pattern	**(File,)(\d+)(items,)(\d+)(pp)**

Pattern replacement	Level of Description:: File Extent and Medium:: $2 $3 $4 $5 Scope and Content::
Input text	File, 2 items, 2pp
Output text	Level of Description:: File Extent and Medium:: 2 items, 2pp Scope and Content::

This pattern matching step is executed at the beginning of the harvesting process. The complete text file is examined and automatically altered when matching patterns are located anywhere throughout the text. On conclusion of this step all values in the texts are now prefixed by a glossary term and can therefore be located and associated with the appropriate term. This also means that values can be processed correctly, in so far as the values associated with different terms will not be concatenated in error.

3.6.3.2. Term Location Calculation

On conclusion of the initial pattern matching step, the next step in the pipeline is to search and locate all glossary terms contained in the text. By reading sequentially through the glossary of terms the program searches for occurrences of each term throughout the whole text file. If the default term is not found in the text the program checks for occurrences of the first alternate term. If this is not found it checks the second alternative term and so on until it finds a match for the fourth alternative term or all possibilities have been exhausted and the term is not found in the document. For each term occurrence found in the text the program stores the term and it's the location (starting position) are inserted into a table.

This process continues until all terms in the glossary have been checked. At the end of this process all term occurrences in the complete document and their starting positions have been captured in a table in the database (Figure 3.22). This table lies at the core of the harvesting process and drives the process until the last term has been processed. It contains the following fields: 'archterm' stores the actual term found in the text, the 'archNewRecYN' field determines whether a new record should be added to the database (picked up automatically from the glossary table). The 'archpos' field denotes the character at which the term starts in the text and the 'archfieldname' contains the name of the destination field into which the value associated with the term will be placed (generated through SQL scripts when the decision to harvest archives was made.) The main processing to extract the key value pairs from the texts may now commence.

archmainref	archterm	archNewRecYN	archpos	archfieldname
IE/NGI/Y33	Reference Code::	Y	177	eadid
IE/NGI/Y33	Title::	N	207	unittitle
IE/NGI/Y33	Dates of Creation::	N	297	unitdate
IE/NGI/Y33	Level of Description::	N	334	unitdesc
IE/NGI/Y33	Extent and Medium::	N	365	extent
IE/NGI/Y33	Name of creators::	N	442	origination
IE/NGI/Y33	Biographical	N	493	bioghist
IE/NGI/Y33	Archival history:	N	6982	custodhist
IE/NGI/Y33	Immediate source of acquisition::	N	7264	acqinfo
IE/NGI/Y33	Scope and Content::	N	7347	scopecontent
IE/NGI/Y33	Appraisal, destruction, and scheduling information...	N	7735	appraisal
IE/NGI/Y33	Accruals::	N	7824	accruals
IE/NGI/Y33	System of arrangement:	N	7856	arrangement
IE/NGI/Y33	Conditions governing access::	N	7948	accessrestrict
IE/NGI/Y33	Conditions governing reproduction::	N	8066	userrestrict
IE/NGI/Y33	Language/scripts of materials::	N	8307	langmaterial

Figure 3.22 Harvesting archives table showing the positions of Terms found in the input texts

3.6.3.3. Extracting key values

This represents the core processing of the harvesting process. At this stage, the pre-processing steps have completed and glossary terms prefix all values in the text file. Additionally, all occurrences of glossary terms used in the document are captured in a table in the database. The purpose of this step is to extract the key terms and their associated values, in preparation for writing to the database. Values are extracted based on the distance between a term and the successive term in the text file until the program reaches the end of the file. The algorithm (Figure 3.23) proceeds as follows. The table of stored terms/positions (Figure 3.22) is sorted in ascending order of position in the text ('archpos' field) and loaded into an array.

The array is processed sequentially and for each term/position in the array the program calculates the first location of the *value* associated with the term. This is equal to the start of the term + the length of the term (thisStartpos). Reading the next element in the array provides the start position of the next term (nextStartpos), therefore the end position of the current value being processed is the nextstartpos – 1. The value can now be extracted as a string of characters between and including the variables thisStartPos and NextStartPos. The extracted value is cleaned before storing to the database.

A new record is added to the database if the flag 'archNewRecYN' ' is set to Y. This should be the case for the term Reference Code, as this indicates a new section of archival materials. Otherwise, the value is stored in an existing record for the associated reference code, into the field 'archfieldname' ' associated with the term being processed.

```
$msearchstring = "";
$startpos = $datastart = $foundpos = $currentpos = $nextpos = $mdatalength = $startposnextterm = 0;
for ($i=0;$i<$arrLen;$i++){
		$msearchstring = $arrFoundTermsPos[$i][1]; // term to search for
		$mfieldname = $arrFoundTermsPos[$i][4]; // name of the field to write to in the archive record
		$startpos = $arrFoundTermsPos[$i][2]; // position of the term in the doc
		$startNewRecord = $arrFoundTermsPos[$i][3]; // set to Y or N
		if ($i+1 <$arrLen){
				$startposnextterm = $arrFoundTermsPos[$i+1][2]; // position of the next term / string in the doc
		}else{
				$startposnextterm = $lastCharPos - 1;
		}
		$foundpos = strpos($filecontent1, $msearchstring,$startpos);
		if ($foundpos > 0) {
				$termlen = strlen($msearchstring);
				$datastart = $foundpos + $termlen; // set start position of DATA value associated with the term
				$mdatalength = $startposnextterm - $datastart;
				$mdata = substr($filecontent1,$datastart,$mdatalength);
				$mdata = udf_tidyContent($mdata, $msearchstring);
				$mdata = udf_discardHeadings($mdata);
				$mdata = udf_trimleadingchars($mdata, ":");
				$mdata = udf_trimleadingchars($mdata, ",");
				if (trim($msearchstring) == "Reference Code::"){
						$m_refcode = $mdata;
				}
				// write a new record if the archnewrecyn (element 3 in the array ) is set to Y
				if ($arrFoundTermsPos[$i][3]==="Y"){
						udf_InsertNewArchRec($p_archDocref, $m_refcode, $parentID);
						$mCurrentRefCode = $m_refcode;
				}
				else
				{
				// update existing record in the DB
				udf_updateDB($p_archDocref, $msearchstring, $mdata, $mfieldname, $mCurrentRefCode);
				}
		}
}
```

Figure 3.23 Algorithm to extract key value pairs and write to the database

3.6.3.4. Value cleansing

Before values are stored in the database a cleaning process is carried out. In addition to removing leading and trailing blank characters as well as leading colons, other strings of data may have been unintentionally appended to some values. The reason for this may be that, an examination of the original texts shows that ISAD(G) Category/area headings are included in the texts presumably for user readability. These extraneous strings would have been appended to the last term/value in category because they were not prefixed by a term. An array comprising the descriptions of the seven Category/area headings is processed sequentially to check for a match in a value. If a match is found then the Category/area heading description is removed from the value. Cleaned values are returned to the Key/value pairs extraction step, or to whatever step initiated the cleaning.

3.6.3.5. Duplicate archives/texts

During the course of testing the system the same texts were input multiple times to the harvesting program. In reality, this may happen. Archivists may amend descriptions of items or update the biographical history of the creating entity for example, as new information is acquired. The 'Accruals' element of ISAD(G) foresees that new materials may be acquired and added to the archives, again highlighting that the archives can change.

The harvesting algorithm automatically attempts to insert a new record when the term 'Reference Code' is processed (because the flag addNewREcYN is set to Y for that term). However, the reference code may already exist in the database if the archives were previously harvested. In that case the system does not add a new record for the reference code. Instead it simply replaces the values already stored in the database with every record now being harvested, for that fonds (collection), without comparing previously stored values with those currently being harvested. This approach guards against duplicating archives and automatically accepts all amendments and accruals (new materials).

The impact of this approach, however, is that it overlooks the fact that expired or removed archival materials may still exist in the database. Whether archives actually expire or are removed is debatable, but, there may be the case of an error in documentation whereby an item is placed in the wrong series or file and is subsequently amended by the archivist. Maintaining the integrity of the archives in the database requires a function to delete records from the database where no corresponding entry exists in the input texts (archives documentation). Controlling which records to remove from the database may be controlled by comparing the time stamp or date last updated on the records for that fonds against the timestamp of the harvesting process whereby all records with an earlier date are removed from the database.

3.7. Research methods and data collection

Section 3.1 discussed the selection of an exploratory mixed methods research (MMR) approach to gain insight into the domain. This section describes the methods and the associated data collection and analysis techniques employed. Adopting the MMR, the researcher elected to conduct an initial case study to improve understanding of the domain and gain insight into how art experts conceptualize their role and the aesthetic objects in their care. Selecting a follow up survey provided a complementary perspective from a wider population and tested the prevalence of the findings from the case study. Finding from this MMR will assist in honing the research question and inform the design of the computational representation in the study.

Of particular interest to this study is the extent to which curators use collateral materials to build knowledge of the aesthetic objects and to describe the objects, and more specifically, the case study explores how primary sources of artist archives contribute to the art specialist knowledge and insight into the authentic natures of the work of art.

3.7.1. Case Study

A case study is an empirical enquiry that investigates a contemporary phenomenon within its natural setting, especially when the boundaries between phenomenon and context may not be clear (Yin, 2014, pp. 14-17). Unlike in laboratory or field experiments, case study researchers typically have less a priori knowledge of the specific variables of interest and they do not seek to control or manipulate variables under study (Ibid.) As a research method, case study is particularly suitable to investigate cases such as decisions, individuals, organizations, programs and processes among other cases, based on hypotheses developed during initial investigations and where "why" and "how" questions seek to understand the nature and complexity of the phenomenon (Ibid.;Creswell, 2014, p. 40) and for the exploration and further hypothesis development stages of the knowledge building process.

A case study fits well to satisfy this researcher's need to gain insight into the case, that being the curator and the important aspects of the curator's world regarding the kinds of things and their relations that art specialists consider when building knowledge of the collections in their care. The researcher is the key instrument for data collection and most of the data were collected in the art gallery or museum. The data collected and the approaches to data collection are discussed below.

3.7.1.1. Data Collection

Guiding principles of data collection for this case study include: Use of multiple data sources; creation of a case study database to hold all evidence collected and the final report; maintain a chain of evidence; and exercising care when using data from electronic sources. Multiple sources of evidence in this study include, interviews with art specialists, observations, physical artefacts, and documentation and strengthen the findings through corroboration of evidence from the different sources (Creswell, 2014; Yin, 2014, pp. 102-130).

Of the many and diverse sources of evidence available to the researcher such as, documentation, archival records, interviews, audio-visual materials, direct observation, participant observation and physical artifacts, Yin (2014, pp. 103-118) claims that the interview is an essential source of evidence. In addition to allowing the researcher to focus directly on the relevant topic through a targeted line of questioning, interviews can yield valuable insight into the matter under investigation, especially when the interviewees are well-informed (Ibid., Creswell, 2014, pp. 190-191). Four well-informed art experts were purposively selected for interview based on their knowledge and work experience at the museum and their academic qualifications. In accordance with the research ethics recommendations (3.1.3) all participants were informed of the purpose of the research and their consent was sought (Appendix F) before the interviews.

Yin (2014) differentiates between a) prolonged interviews with a duration of two or more hours in one sitting or extended over a period of time where the interviewer seeks the opinions, interpretations and insights into particular occurrences and b) shorter, more focused interviews of approximately one hour where the purpose may be to substantiate initial findings through open-ended questions that likely follow the interview protocol more closely.

The research literature also classifies interviews by structure:

a) Unstructured or intensive interviews require intensive planning for any contingency whereby the interviewer employs open questions and develops a list of alternative reactions as the conversation develops. In contrast to other interview structures, questions are often determined while the data are being collected and this type of interview is more appropriate for experienced interviewers.

b) Semi-structured interviews employ a predefined protocol, enabling comparison between interviewee responses. However, flexibility in question sequence and incorporation of unstructured questions and more in-depth exploration is common (Yin, 2014, pp. 110-113)

c) Structured interviews, similar to surveys, comprise closed questions with predefined responses. Close adherence to the interview protocol, posing the same questions to all interviewees and in the same sequence, leads to standardization and easier comparison between different interviews. (Bryman, 2012, pp. 209-229; Yin, 2014, pp. 112-113)

Observation

According to (Yin, 2014, pp. 106-115) the advantages or strength of observation lies in the contextual nature of the data, which may relate to the behavior of individuals or to the evidence collected from the work environment. Although this can be time consuming and the preferred method is to include more than one observer, it was not possible to have a second observer present and observation was carried out solely by the researcher. Further insight into the domain and the curator's role and conceptualization of art, emerged from: 1) A large part of the art specialist's role involves exhibition curation and observations at exhibitions in different museums, facilitated data collection in relation to the kinds of data presented in collateral texts selected; 2) An opportunity for direct observation presented when the researcher attended a conference with the curator/archivist, whose presentation concerned aspects of documenting materials in the archives, and this highlighted the complexity involved in documenting archives where one object may be attached or embedded in another object; and 3) during one of the visits to the archives the researcher observed the research curator in the mode of 'a gallery guide' and noted her expertise and knowledge of the art work as she described a particular work of art to a notable visitor.

Documentation

Analysis of documentation, comprising metadata standards and vocabularies developed specifically for the domain over many decades and exhibition catalogs, provided corroboratory evidence of important data for curators in relation to describing a work of art and collateral materials, some of which were evidenced through other sources (observation, interviews and physical artifact). The advantages of examining these materials included repeated accessibility, they were created specifically for the domain of interest and by experts in the field (Yin, 2014, pp. 105-107)

Physical artefacts

Examination of the paintings, displayed at exhibition revealed material facts of interest to domain experts (by way of example, size, technique, medium, signature), which form part of the tombstone data about the objects. Similarly, examination of artefacts in the archives provided contextual information about the artist and his methods as collateral to the paintings in the exhibition. Art experts may consider these collateral data important to gain insight into the artist's paintings.

3.7.1.2. Trustworthiness and Data Analysis

Yin (2014, pp. 45-49) distinguishes four criteria to measure the quality of case study research: Construct Validity, Internal validity, External validity and Reliability. Construct validity refers to the accuracy with which the case study measures reflect the concepts being studied, and therefore the researcher focused on collecting relevant data using appropriate multiple instruments, as described above to answer the question or to test the hypothesis: (Refer also to *Triangulation* below). Discussions on these matters are in the section above.

Internal validity refers to how the researcher deduces causal relationships between variables and is relevant to explanatory case studies but is not relevant to exploratory and descriptive studies.

External validity in relation to a case study involves seeking generalisable findings to the initial hypothesis or lessons learned, that is *analytic generalizations*, unlike a survey which seeks statistical generalization over a population. Reliability concerns the consistency and repeatability of the study procedures, which relies on the case study documentation and this relates back to one of the four guiding principles of case study research, of establishing a database for all data collections protocols and data, including field notes, observations, interviews and so forth.

According to Creswell (2014) good qualitative research includes researcher comments on how their interpretation of findings is influenced by their background and culture and to avoid bias and to clarify any bias a researcher may bring to a study. Similarly, Yin (2014) emphasises that the researcher must avoid bias and to remain open to evidence contrary to any preconceived position the researcher may have.

3.7.1.2.1. Triangulation

The concept of converging lines of enquiry, or corroboration to build confidence in research findings is considered by many authors (Creswell, 2014; Sekaran & Bougie, 2016; Yin, 2014, pp. 118-123) Creswell maintains that triangulation through multiple sources strengthens both internal validity and reliability. Four types of Triangulation are considered:

1. Methods: using multiple methods of data collection and analysis (e.g. qualitative-case study and quantitative - survey)
2. Data: collection from different sources or at different time (interviews, observation)
3. Researcher: multiple researchers collect and/or analyse the data
4. Theory triangulation: multiple theories and/or perspectives are used to interpret and explain the data

This study employed multiple methods of data collection and analysis (a case study and a survey) and data was collected from multiple sources also within the case study, and different perspectives on the case study findings were provided from the survey findings. However, there was only one researcher involved throughout the study.

3.7.1.3. Data Preparation & Analysis

Prior to the interviews the researcher drew up an initial Code Book (Appendix B 7.1.3) to analyse all qualitative data collected. The codes used were based on the questions on the interview protocol. Assigning codes to the data provides a means to search and sort the data during data analysis(Bazeley, 2013, pp. 125-156). The researcher recorded all of the observation, fieldwork and documents reviewed in both Excel spreadsheets and Word documents and placed these in a Data Collection folder. Each interview was processed shortly after it took place. Listening to the recorded interviews a few times provided a general sense of the data, before transcribing the interview first into a word document and from there into an Excel spreadsheet. Allocating each response to a cell meant that a code could be added to each sentence and if more than one code is needed then the code is placed on a new line as advised by Bazeley (2013)

Directed and indirect content methods were used on the qualitative data (Bazeley, 2013, pp. 125-186; Hseih & Shannon, 2005) to test initial hypotheses and to uncover concepts and themes emerging from the data, and quantitative data analysis used statistical methods. Furthermore, triangulating the data from all sources in the both methods determined the level of consistency or corroboration of the findings (Sekaran & Bougie, 2016, p. 106; Yin, 2014, pp. 120-121)

Due to the costs of software systems to analyse qualitative data, such as, NVivo10, the researcher instead used excel spreadsheets and used colour codes to assist in the analysis of the data (Appendix B 7.1.2). Each interview separately and then combined all four spreadsheet into one large spreadsheet and sorted the data by the code.

3.7.2. Survey

Choosing to conduct a web based survey provided access to a larger population in a more economic manner and it facilitated a more rapid turnaround in data collection than conducting interviews. On the other hand, careful questionnaire design and pre-testing was imperative, to remove ambiguity to promote consistency of interpretation across the population. Designing the survey involved consideration of the purpose of the survey, choice of whether to use a probability sample, the sample, frame, the sample size, questionnaire design and response rates (Creswell, 2014, pp. 155-159; Fowler Jr, 2014, pp. 6-11). As part of a mixed methods study the survey questions were the same as those asked of participants in the case study. The purpose of the survey was again to explore the role and information requirements of domain experts, but this time, across a larger population of domain experts. The survey also tested the prevalence of the findings from the case study among this population and to gain an alternative perspective on those findings. Therefore, the survey was not necessarily designed to determine probability of findings to generalise to a population (Ibid p.14). In an effort to limit sampling errors the sample frame was restricted to curators and domain experts who carry out research on works of art and the sampling procedure was random in so far as the researcher randomly browsed museum websites trawling for email addresses of the curators listed on their websites. Sample stratification did not occur before or during sample selection but optional questions relating to demographics were included on the final page of the survey to enable analysis and reporting by country, gender, age and qualifications if required.

3.8. The role of properties

This section describes the role of properties in the domain, encompassing the definition of specific aspects such as the characteristics and attributes of objects, establishing and describing relationships, supporting search and retrieval and connecting with external objects.

First, properties describe the attributes or characteristics of interest about entities in the domain, where each named property maps to a value set or domain that restricts the values and data type of the property. By way of example, the value mapped to the property ‘Date Created’ restricts the value to a signed integer data type, representing a year; whereas the property ‘Date Updated’ maps to a Timestamp and the property ‘Artwork Description’ may be limited to a string of 2000 alphanumeric characters. Classes/Entity types such as Agent, Object, and Archival Object can then be described by relating or combining a number of properties into a tuple/record, as appropriate to each class.

However, in relation to the identity property, or distinguishing specific instances from other instances of the same class, requires some property which is unique within the entire entity type/class. In the relational database model (RDB), Primary key constraints fulfills the identity property and comprises either one or more of the objects’ properties or employs a surrogate key that maps to a value set that ensures uniqueness. By way of example, assigning an Agent

ID: 500002320, distinguishes the artist Jack B Yeats, from Agent ID: 500016889, John Butler Yeats. Thereafter, the system refers to instances by the unique identifier.

Apart from describing the characteristics and related data types and establishing the identity property, properties also define and participate in relations between instances in different entity types/classes and also among instances of the same entity type. The Foreign key constraint of the RDB enforces a relationship between instances of different entity types, in conjunction with a unique index or Primary key. Alternatively, relations between entities in the same entity type/class can be achieved by storing the identifier of the related entity, in a property ParentID, for example, an object in the archives identifies the parent archival record (perhaps a file or series record), to which it belongs. These examples, however, restrict the degree of the relations to a one-to-one (based on the foreign key) or one-to-many in the latter relationship.

On the other hand, combining properties from the same or different entity types defines multi-valued *relationship* entity types. By way of example, an object such as a sketch or a letter from the archives, may be related to multiple artworks and vice versa, each artwork may be related to many objects in the archives. These relations are realized by storing the primary key properties from both the artwork object and the archival object in a separate *relationship* entity type and this also automatically establishes reciprocal relations between the objects. Similarly, combining the primary key properties of an artwork object and subjects/keywords into a relationship entity type (Table 5.3) facilitates relationships between artworks and subjects/keywords. Enabling relationships among different Agents and between Agents and Events (Table 5.2) also defines new relationship entity types, based on properties of primary keys, both from Agents in the former case and from the Agent and Event entities in the latter case. Furthermore, additional properties defined on a relationship provide supplementary information about the relationship, as evidenced when explicitly establishing relationships between archives and artworks and allowing the art specialist chooses a term to describe the relationship and to record his or her reflections about that relationship (Figure 5.12).

Properties defined on the generalization/superclass are inherited by the subclass. By way of example, the painting entity type automatically inherits all properties from the Objects class and additional properties are defined on the painting entity.

In addition to describing and relating objects in the domain, particular object properties facilitate search and retrieval. To enhance Search performance indexes may be created based on a number of properties, such as Acquisition number, Artwork Title, Year and Artist. Similarly, the archival objects' properties facilitate searching by date range, object type, record description (albeit the description remains at the high level of archival record, due to lack of properties describing various aspects of individual sketches, for example). Alternatively, some properties facilitate wildcard searches of terms in the title and descriptions of both the artworks and archival objects. Other properties, for which indexes may not exist, may still be included in a search but would be considerably slower if the collections contained large volumes of objects.

Finally, a property may provide access to external resources. By way of example, a property of the Agent entity type describes the URL of a ULAN record and facilitates access to biographical and bibliographical materials about the artist.

3.9. Conclusion

This chapter discussed the methods adopted for this study, including conceptual modeling, logical modeling and harvesting archives from source texts. The discussion of relational calculus and related technologies together with the widespread use of relational databases has convinced this researcher to use this robust, scalable and proven technology. Furthermore, the relational database model supports the rich representation of objects in the domain and the interconnectivity of aesthetic objects and their collateral data. The researcher acknowledges the existence of alternative technologies, some of which (for example, OWL, Topic Maps and OAI-PMH) are discussed in the literature review. Recall however, that the majority of existing systems implemented in the cultural heritage domain are founded upon relational database technologies (Table 1.1) and archives and artworks systems are implemented as separate systems in many museums. Moreover, this study is concerned with extending existing systems and integrating with existing systems to incorporate archives with artworks in one representation. Perhaps superimposing a Topic Map over these base systems and potentially referencing additional systems that exist in museums is a viable option for the immediate future rather than compelling the re-engineering of existing systems in Ontologies, Triplestores or Graph databases.

The work involved in designing a front end application to access the database and the development of the databases in this study is described (Chapter 5.) The challenge at hand is to incorporate the knowledge of art specialists responsible for the preservation of aesthetic objects held in collections in their care, in a representation that organizes and describes these objects and their collaterals. To this end a case study will afford the opportunity to gain insight into the domain and to elicit the important aspects for the representation. A follow up survey of the wider art specialist community will provide complementary data and test the prevalence of the case study findings (Chapter 4.).

4. Potential Art specialist requirements

4.1. Introduction

Designing a computational representation that incorporates the knowledge of domain experts responsible for the preservation of aesthetic objects in museum collections, presents an interesting challenge; especially considering the views expressed in the philosophical literature (Lamarque, 2012; Margolis, 1977; Thomasson, 2004; Walton, 1970) highlighting the complex nature of these objects.

The literature review also shows, however that an immense body of knowledge exists, embedded in tools and standards developed specifically for the domain, to organise and describe these complex objects. While these domain tools and standards will be valuable resources for this research, knowing how best to select, evaluate and to use these tools to accurately represent essential aspects of the domain, needs further investigation. Moreover, the researcher requires not only an understanding of the semantics of the domain, but of equal importance, the researcher needs to understand the important aspects of the domain, from the perspective of the curator.

Given the complex nature of these objects and combined with the fact that the researcher has limited knowledge of art and curatorial practices, the need for access to the knowledge of art experts responsible for the preservation of these collections, was clear.

This chapter details the procedures and the steps carried out, to acquire knowledge about the domain and to gain insight into particular aspects of the domain from the art specialist's perspective. First, the context in which this research is carried out, is described (4.1.1). Next, employing an exploratory sequential mixed methods research approach fits well with this thesis, providing a framework that assists in establishing a sound theoretical and technical foundation for the computational representation (4.1.2). A case study that collects and analyses data from interviews with 4 domain experts (among other data sources) and the findings from the case study, are presented (). Findings from the case study serve as input to the design of a questionnaire, and survey design, implementation, analysis and finding based on 48 valid responses from domain specialists is described (4.3). The chapter concludes with a discussion of the findings (4.4).

4.1.1. Context

Typically, metropolitan cities in the world comprise museums and galleries that preserve and transmit their cultures on the basis of aesthetic objects. Within Dublin there are five main museums: the National Gallery of Ireland (NGI), the world renowned Chester Beatty museum, Dublin City Gallery the Hugh Lane, the Irish Museum of Modern Art and the Douglas Hyde Gallery at Trinity College. In a sense, the national galleries and museums comprise a representational set of the national culture. One example of aesthetic objects in this representational set is painting.

The National Gallery of Ireland (NGI) houses a collection of over 16,300 artworks, spanning the history of Western European art from about 1300 to the present day, while it also features the world's most comprehensive collection of Irish art[38]. The gallery holds exhibitions of paintings on a thematic basis, to celebrate one or more artists or paintings or to promote some cultural aspect of the nation. Paintings are displayed showing details such as title, short description, credit line and other attributes and all of these details are stored in a database. However, a very important function of NGI relates to preservation, including those involved in physical preservation and those involved in preservation of the intellectual value of the collections and carry out research of a work of art, and facilitate investigation into the painting. This investigation is facilitated by means of historical records for example, objects related to paintings, art criticism texts and even the physical legacy of paintings such as paint brushes, clothing and so forth. These collateral objects related to one or more paintings have to be catalogued, stored, retrieved and annotated by curators.

NGI also holds a major archival collection; the Yeats archives. The Yeats archive was donated to NGI in 1996 by Ann Yeats (daughter of the poet William Butler Yeats who was the brother of the artist Jack B Yeats) with additional collections acquired since then. This large and eclectic collection of materials of Jack B Yeats and the Yeats family 'paints a picture' of the life and times of the artist and provides insight into the artist methods The collection comprises items such as watercolours; sketch books; books and plays authored by the artist and also collaborations with others (for example, the playwright J M Synge, the poet W B Yeats); Broadsheet and Broadside publications along with physical legacy of his work such as the artist smock; paintbrush; glove; and easel and memorabilia including a miniature stage built by the artist which he used to stage plays entertaining children in the neighborhood when living in the UK.

The focus of this study is on aesthetic objects, such as paintings, more specifically, this research considers the paintings of Irish artist Jack B Yeats and the Yeats archives at NGI represent potentially rich resources and collateral materials that will be of great value to this study. Access to art experts at NGI will provide the specialist knowledge of curators involved in the preservation of these collections.

Having considered the above museums and the focus of our study, we approached experts at the National Gallery of Ireland (NGI). Consequently, this research is carried out in cooperation with experts at NGI, in that, it is informed by inputs from curators, an archivist, a collections management system administrator and a research curator, over many interactions with these experts.

[38] https://www.nationalgallery.ie/art-and-artists/painting-and-sculpture, www.nationalgallery.ie/art-and-artists revisited 9/03/2018

Also, including a curator from the Chester Beatty Library (CBL), a designated National Cultural Institution [39], offers a complementary perspective on the domain since the CBL collections of Western, Islamic and Asian art, represent cultures and religions of the world, ranging from about 2700 BC to the present day.

4.1.2. Aims of this mixed methods study

The ultimate aim of this phase of the research is to acquire the relevant knowledge about the domain, specifically from the perspective of domain specialists, to inform the design of a ‘good’ computational representation that incorporates the knowledge of curators. Since a mixed methods research (MMR) approach enables stronger evidence, through complementary methods and data collection and analysis, based on the same questions (Creswell, 2014, pp. 217-219; Venkatesh et al., 2013; Yin, 2014, pp. 65-67), rather than using one method alone, this study comprises a case study followed by a survey. The case study explores the role of the curator and how these art specialist approach a work of art, the kinds of information resources they use to develop knowledge of their collections and it explores a number of resources developed for the domain. Findings from the case study feed into the design of a survey, which tests the prevalence of finding from the case study and engenders deeper insight into the domain through complementary (quantitative data) perspectives, from a larger population. In a sense, the MMR serves a similar role as the requirements gathering and analysis phase undertaken by computer science and information science practitioners involved in systems design. In this case the researcher seeks to enhance understanding of relevant aspects of the domain, so that, ultimately, the best or most appropriate design choices are made, for the computational representation. The case study and survey were carried out in compliance with ethical standards as set out by the Research Ethics Committee (3.1.3)

4.2. Case Study

As an empirical enquiry that investigates the nature and complexity of a contemporary phenomenon within its natural setting (Creswell, 2014, p. 40; Yin, 2014, pp. 14-17), this case study explores the ontological commitments of curators, responsible for the preservation of collections in their care. The majority of the data was collected in the workplace; either in the curators’ offices or in the gallery and museums at exhibitions, at meetings with domain specialists and searching the archives with domain specialists. Examination of standards and documentation took place in the office at Trinity College Dublin.

4.2.1. Purpose

The purpose of the case study is to gain insight into the curators’ conceptualisation of their role, of works of art, how they approach a work of art, and the kinds of things and their relations that curators consider when researching objects in the collections in their care. In particular, and in light of the increasing interest in artists’ archives as potentially valuable resources for art

[39] http://cbl.ie/Collections/Introduction.aspx last visited 10/01/2018

research (Breakell, 2015; Koot & Koppenol, 2016; Pyle, 2001; M. Reed, 2017; Sjöholm, 2014) this study explores how and to what extent curators feel that this is the case.

In seeking to understand the domain from the curator's perspective, interviews with art specialists focused on the following topics:

- The role of the curator
- How primary sources of artists archives contribute to art research
- Information about works of art that curators consider important
- The information sources considered relevant to scholarly research of art

Since the overall objective of this thesis is to design a computational representation, the study also includes exploration of art materials, fieldwork at exhibitions and it explores standards developed within the domain, to determine how they facilitate capturing domain expert's knowledge.

Findings from this study will enhance the researcher's understanding of the domain and will assist in refining the research question (Rai, 2017; Yin, 2014, p. 238). Moreover, the findings will inform the design of a questionnaire for a follow up survey that will provide improved measurement over a wider and larger population of art specialists (Creswell, 2014, pp. 225-227).

4.2.2. Data sources

Interviews with 4 domain experts provide insight into important aspects of the reality of the domain, from the domain experts' perspective. In accordance with the principle of using multiple sources of evidence (Yin, 2014, pp. 103-130) and also employing sound knowledge engineering practices, such as, 'the heuristic of specific location' (Winston, 1992, pp. 168-169) the researcher enhances the quality and validity of this study by exploring and collecting data from a number of different sources. Data collection from fieldwork, including visits to exhibitions and review of exhibition catalogues and art materials provides further evidence and insight into the kinds of information they consider important for art research. Exploration of domain standards not only provides information about important aspects of the domain from a more general and technical point of view, which feeds into the design of the computational representation, but it examines how these tools serve the needs of domain experts and how they incorporate domain experts knowledge.

4.2.3. Sample Frame & Sample Size

The sample frame is restricted to domain specialists involved in the preservation of aesthetic objects in museum collection and those who carry out research of works of art. The sample was purposively selected, in that, we approached art specialists at the National Gallery of Ireland (NGI) and a curator at the Chester Beatty Library (CBL), who volunteered to participate in this study (Table 4.1).

Table 4.1 Profile of art experts interviewed in this study.

Inst.	Expert Role	Education	Education Level	Years Experience at the gallery
NGI	Head of Research & Collections	History of Art	PhD - 2004	13
NGI	Exhibition Curator & Manager CSIA – Archives	Multimedia & History of Art	Masters - 2012	12
NGI	Research Fellow (Specialising in Irish art 1850-1950)	History of Art	PhD - 2014	8
CBL	Curator of Western Collections	History of Art & Architecture	PhD - 2010	12

At the time of the interview expert A's official title was Curator of Northern European Art, at NGI, with responsibility for the Dutch, Flemish, German and Netherlandish collections. Specialising in Northern European art, he focuses on Dutch art, especially on Dutch genre painting, including artists such as Gabriel Metsu and Vermeer. He holds a PhD in the History of Art, from New York University and prior to joining NGI in 2004 he worked at the National Gallery of Art in Washington, briefly. Subsequently, he was appointed as Head Curator in 2016 and as Head of Collections and Research in June 2017.

The second expert manages the ESB Centre for the Study of Irish Art (CSIA), a research facility at NGI, housing the gallery's research collections of archival and library materials associated with the history of Irish Art. The CSIA holds two categories of materials: a primary collection of unique materials of artists (including the Yeats archives) and arts organisation papers; and a secondary collection comprising mainly ephemeral materials, published by galleries and publishers over the last one hundred and fifty years. Previously employed in the education department, from 2001, expert B' was involved in digitisation projects of the library and archives and early versions of the NGI web site. More recently, he now curates an exhibition each year at the gallery. He holds a degree in Fine Art painting and History of art, a Masters in Multimedia and a Master of Philosophy in the History of Art.

Expert C holds a two year research position at NGI, supporting exhibitions, publications, visiting researchers and working on other internal projects, as they arise. Prior to commencing the Research Curator role at NGI in early 2015 and similar to expert B, she also has a longstanding relationship of about eight years, with the gallery. Following an internship in the Education department at NGI which she completed during her academic studies, she continues to provide guided tours for primary and secondary schools and the general public, in the gallery. She has degree in Art History and English, a Masters in Irish Art History and a PhD in Art History, from Trinity College Dublin, completed in 2014.

As curator of the Western collection at the Chester Beatty Library (CBL), expert D is responsible for the Western collection, comprising primarily Christian materials but also includes non European Christian materials, for example, Ethiopian, Armenian, Egyptian and Middle Eastern Christian Manuscripts. Additional materials include anything European that is non Christian, such as Egyptian papyrus materials and a limited but significant early collection of cuneiform tablets. Prior to her appointment as Curator in 2013, expert D's longstanding relationship with CBL extends from 2005, in the role of researcher and curatorial assistant to the Islamic curator. She also worked short-term and then part-time while completing a PhD in medieval architecture at Trinity College Dublin, in 2010.

4.2.4. Pre-Interview testing

The researcher designed an interview protocol and conducted pre-interview tests with two people prior to interviewing the domain experts. The first interviewee holds a Degree in Fine Art and a Masters in Interactive Digital Media (2013), from Trinity College Dublin. The second person holds a Degree in Fine Art from Moscow University, a Masters in Fine Art in the Digital World from NCAD (National College of Art & Design, in Dublin) and a Masters in Computer Science (2016) from Trinity College Dublin. The purpose of these interviews was twofold: question refinement and interviewer preparedness. Feedback from interviewees resulted in rewording some of the questions to remove ambiguity, jargon and irrelevant questions. Furthermore, the practice sessions clarified the procedure for the researcher and identified areas for improvement of researcher questioning and listening skills. Also, the practice sessions uncovered timing issues and it allowed testing the voice recording apparatus. Although acknowledging similarities between some questions in relation to the resources used, frequency of use, value of the resources, benefits or shortfalls identified, these remained in the interview guide given that perhaps repetition and or approaching the question from a different perspective may assist recall or perhaps reveal new or varying opinions.

4.2.5. Conducting the Interviews

Subsequent to receipt of ethics approval from the research board of Trinity College Dublin, the process of organizing on-site interviews commenced in April 2016. Prior to each interview, the researcher emailed the study Information page, Participant Informed Consent Forms and the Interview Guide (Appendix A) to each interviewee. Providing an indication of the nature of the study, participant criteria and rights, the upcoming questions and the kinds of information relevant to the study, afforded an opportunity for the curators to familiarise themselves with the study and the interview questions.

Commencing on 20th April 2016 with expert A, the interviews continued with expert B on 20th May 2016, with expert C on 9th December 2016 and concluded on 23rd January 2017 with expert D. All interviews took place on-site at the domain expert's office, either in the National Gallery of Ireland (NGI) or the Chester Beatty Library (CBL). At the outset, both the researcher and interviewee signed the Informed Consent forms and on confirmation that the experts had read the forms beforehand, the researcher briefly clarified the purpose of the interview and participant criteria and rights and tested the recording equipment.

4.2.6. Data Collection

4.2.6.1. Meetings

A series of meetings took place with domain specialists, including curators, archivists, a librarian and an administrator of the collections management system, over the course of the four years of this study. Commencing in October 2013, the meetings continued until May 2016 (Appendix D) and a summary of the principal topics discussed, follows.

Curator of the collection: Important tasks for the curator involve long term planning and research into works of art, and we identified two primary tasks: a) exhibition and b) research and publication. The curator informs that an exhibition may celebrate a particular event; may be organized around a particular theme; celebrate a group of artists or an exhibition may be organized on some other basis that is of interest to the public. Furthermore, curating an exhibition can take a couple of years, from inception through to actualization and sometimes galleries work in cooperation with each other, agreeing the loan of an artwork for the duration of the exhibition. Discussions about the curators recently published book provided insight into how the art specialist conceptualizes a painting and one observes that the curator considers many aspects as well as and beyond the “image” important for research in art. In developing knowledge of a painting, the art specialist takes into account details such as, the artist life; the artist’s oeuvre; the artwork in relation to the other artworks; the social and cultural context in which an artwork is created; artwork attribution and provenance. Furthermore, the curator identified the main subject or theme for each painting and classified these into groups, in this instance as “Search Aids”.

Curator of the archives: Discussions with the curator revealed the fundamental guiding principles for archivists; “respect des fonds” meaning that records in the archives are documented in the original order, as originally organized by the entity that created the archive thus preserving authenticity and context of the records. The documentation standards employed at NGI for archives include the General International Standard Archival Description (ISADG).

Many visits to the Yeats archives (See 4.1.1 for a description of the Yeats archives) with the archivist and a research curator provided the opportunity to examine the materials and to get a real sense of the artist and his life and his working methods, and for this researcher an increasing appreciation of the artist and the value of these materials for art research. During these visits the researcher took photographs of the materials (to include in the prototype) and signed a form agreeing to NGI restrictions on the use of the materials.

At one of the earlier visits to NGI we had the rare opportunity of viewing a recently acquired collection: the personal library and archives of Sir Denis Mahon (collector and renowned specialist in Italian art), whose vast collection was housed in very large, mobile shelving, awaiting curation. This visit provided further insight into the role of the curator or archivist of these materials. The archivist visited Sir Denis before his death, on a quest to acquire the archives and the successful acquisition of this prestigious collection is recognised as a major achievement for NGI. We also learned more about the methods and tasks of the archivist, comprising, collection inspection and conservation; establishing an inventory of each item; researching and documenting each item; collaborating with specialists in preservation; dealing with the public and researchers seeking information from the archives; presenting materials at conferences and organizing tasks and workloads of new assistants. Finally, we learn how the curator considers the possible sensitive nature of archival materials and exercises discretion in releasing materials for public access, even if the thirty year restriction has elapsed.

Computer Systems: A number of computer systems are used at NGI, and similar to other institutions' websites, NGI's website informs the public about exhibitions, visiting times, learning, research and the museum shop. However, of particular interest to this study, is the fact that the website provides public access to the museum's collections. The front-end application to the archives "Doras", provides access to the backend archive system "iBase", a digital asset management system, but due to copy-write restrictions, the Yeats archives are available online in MS Word document format only. 'Portico', the front-end application to the library may be accessed from the website. eMuseum is the front-end application to the museum collection. eMuseum, provided by Gallery System, publishes / retrieves works from the collection. The back-end collections management system is TMS (the Museum System, based on a relational database), also supplied by Gallery Systems. A variety of external systems used by gallery personnel, include frequently accessed systems such as; Art Index (an online database providing access to scholarly journals and other art specialist publications) and Auction house systems (subscriptions required for both systems) used to track provenance and movements of works of art.

Responding to a query with regards to minimal information published on the website, for example, a description of the artwork or details of the artist is not available, the librarian informed us that in earlier years a description would have been compiled, but that now the practice of describing an artwork is not always carried out, rather the curator keeps a (paper based) dossier for each artwork.

4.2.6.2. Semi-structured Interviews

Interviews with the domain specialists commenced in line with the interview protocol, but given the nature of semi-structured interviews, it facilitates exploration of other relevant topics as the opportunity arises. Asking questions about their role, and kinds of things of interest to the art specialist when researching a work of art, and about the resources that they use, allowed the researcher to explore the extent to which domain experts adopt 'contextualist theory', (that proposes that information external to an artwork is crucial for appreciation of the work) or if the alternative 'formalist or aesthetic empiricist theory' (which holds that all necessary information for appreciation of an artwork is intrinsic to the work), prevails. In particular, the interview explored expert opinions on the significance of artists' archives for their research of a work of art.

Depending on the experience of the curator, questions differed at various points throughout the interviews, where, for example, one of the experts had experience in provenance research, whereas two of the experts had more experience in the cataloguing systems at their institution. However, all interviews attended to the main themes and some of the interviews included parallel or relevant topics that arose in conversation. Each interview was recorded and lasted approximately 55 minutes.

4.2.6.3. Fieldwork – data investigation at exhibitions

The researcher visited a number of museums during the course of this study, including: "Masterpieces from the Collection, 2013"; the annual exhibition of Turner watercolours, each January at NGI, Vermeer in June 2017 at NGI, visits to Chester Beatty Library (CBL), Hong Ling exhibition at CBL, Tate Britain and Tate Modern in June 2016, Magritte gallery in Brussels in 2014). Through these visits, the researcher gained insight into the role of the curator and the knowledge that curators possess about the works of art in their care. Observations during these visits contribute to an initial catalogue of facts relating to works of art (Table 4.2).

Visually perceptible data constituting the physical artwork noted include items as shape; dimension; technique; and supporting medium. Techniques, for example, Oil, Pastel and supporting medium such as board and canvas indicate that a variety of both types of data may exist. Also, the artworks can be classified as type painting.

Beside or under each artwork, a label lists the material facts (dimension, technique and medium) along with external information about the artwork and the artist, including: title; artist's name; birth year and location and the year and location in which the artist died. Flags, such as c. (circa) are used to denote 'around the Year specified' and a question mark denotes that the year specified is questionable. A caption or brief textual description, provided for some of the artworks only, may focus on some element depicted, relating that to the theme of the artwork and associated historical context. At the National Gallery of Ireland, the Irish language version featured at the top of the label, but the English language translation also appears underneath.

The Credit Line of the artwork, by way of example, "Purchased 1965 (Shaw Fund)" is displayed for only some of the paintings and additional acquisition methods observed include: Bequeathed; Purchased; Donated; On Loan. A unique identifier for the artwork is displayed on the label, but the format of the identifier differs across artworks. The majority are in the format NGI followed by a number, while examples of other formats observed are L.2009.1, 2011.13, denoting the artwork is on loan (L.2009.1) to the gallery and the new acquisition coding structure adopted by the NGI, as described (in the next section) by the curator, respectively. The Chester Beatty Library displays text panels, describing printing techniques such as

Etching, Engraving, Woodcuts and panels on book design, including Ancient books, Illuminated books and Book design and the Graphic Arts, next to each matching group of exhibits.

Table 4.2 Metadata relating to work of art (observed during visits to exhibitions)

Of the Object **Visually perceptible data**	**About the Object & Related to the object**	**Acquisition Methods**
Physical/constituents	*Label* containing:	Gift
Shape: rectangular	Title	Purchase
Dimensions: Height x Width	Artist Name	Bequest
Measurement Type	Artist Life span - Birth Year & Year of Death	On Loan
Technique e.g. Oil	Birth Location	Donated
Supporting Medium e.g. Canvas	Year qualifier: c. ?, [ ] denoting Circa, Uncertainty, Within a time span	Private Purchase
	Location where artist died	Presented
Content -of/in the artwork:	Caption: (for some paintings) identify Context, theme	
People, buildings, nature, objects etc	Credit Line e.g. Purchased 1965 (Shaw Fund), On loan, Presented, Trustees of the Haverty Trust, 1931	
Art Elements: e.g. line, shape, colour, texture	Unique Identifier: e.g. L.2009.1, 2011.13, NGI.941	
Inscription:	*Text Panel (CBL)* describing:	
Position of Inscription	Technique for a selection of exhibits, e.g. Etching,	
	Types of objects, e.g. Ancient books, Illuminated books, Book design, Graphic Arts	
	Language:	
	English language used to describe the work	
	Caption in Irish language for some objects, in NGI	
	Related Materials:	
	Sketch: of Turner painting next to the painting (NGI)	
	Photographs & Letters: Displayed in cabinets in exhibition hall of René Magritte paintings (Magritte Museum, Brussels)	
	Exhibition	
	Name, Location, Time span, exhibits	

Compiling a list of the artworks exhibited in the Irish section of the “Masterpieces” exhibition, in tabular format (7.1.7) highlights anomalies in the ‘Year’ data and the different formats of the identifiers (recorded for Yeats artwork only) are also clearly visible. Observations (Appendix E) contain notes made during a gallery visit and questions arising during follow up analysis.

4.2.6.4. Documents - Exhibition Catalogue & Art Books:

Examination of the exhibition catalogue entitled “Masquerade & Spectacle: The Circus and the Travelling Fair in the work of Jack B. Yeats” (Kennedy, 2007) reveals metadata associated with aesthetic objects. An initial set of metadata items include: textual descriptions for each painting describing the content and theme, and providing the historical and social context for some of the paintings. Thus, these descriptive metadata elements contribute to the catalog of items describing aesthetic objects, along with the author, date and source material. The content and theme noted in the text descriptions above identifies a metadata element ‘keyword’ to store this information, possibly for future search and retrieval. The catalogue also provides a succinct list of major events in the life of the artist and capturing these events in metadata elements also contributes to the representation of aesthetic objects and contributes to the intellectual value of the objects. Further exploration of the exhibition catalogue underscores several important curatorial tasks, associated with the curation of an exhibition, comprising: a) devising a title that encapsulates the primary theme of the exhibition; b) composing a general description of the exhibition; c) selecting the artworks and d) arranging the artworks and e) organizing dates and location for the exhibition, and all of these details contribute to the entity ‘Exhibition’.

Gallery publications include details of acquisitions (NGI printed publications are available in Trinity library) showing a unique code or identifier allocated to each item in the collection. Examination of NGI yearly acquisition publications suggested a pattern in object identifiers, for example, all marble sculptures were allocated a number within a specific range and likewise for paintings. The curator confirms that although this was the practice, that now, each acquisition is allocated the next available number, irrespective of the category of artwork.

A Review of Art Books also highlights metadata for describing an artwork. (Waiboer, 2012) organises artworks into specific categories based on content depicted. In addition to the standard tombstone details for each work of art, further evidence considered relevant to the research, uncovered, in the Plate descriptions of a catalog Raisonné, includes information, such as, current location; Verso; Graphic Reproductions; Copies; Untraced Works; Collection Catalogues; Note and Literature (Ibid.)

Pyle (1992) includes metadata (already identified in the previous section) for most of the c. 1200 oil paintings by Jack B Yeats. She also includes a comprehensive list of past exhibitions of the artist’s work. Evidently, recording and tracking each named exhibition, requires capturing both the Date and Location metadata elements, since the same exhibition (title) may be presented at different times in the same or different locations. Pyle also organises the artworks into high level groups denoting subject matter and associates a number of specific subject *keywords*, with each one. Rosenthal (2003) builds on the work carried out by Pyle; for instance, he lists later exhibitions; presents a family tree of Yeats ancestry and the contribution to the arts by family members, and he also lists significant events in the life of the artist. Capturing these relationships and events in metadata elements also contribute to the representation of aesthetic objects. As previously noted, Jack B Yeats authored books and plays and collaborated with other artists, for instance, with the playwright John B Synge (Pyle, 1989), which is documented in the Yeats archive. Additionally, a metadata element defining the object type (book, play, manuscript and so forth) may be required.

4.2.6.5. Exploration of metadata standards and vocabularies used in the domain

Resources for organizing and describing museum collections comprise (among others) metadata schemata and vocabularies. The following metadata schemata are particularly relevant to this study: Category for the Description of Works of Art (CDWA) (Baca & Harpring, 2014 -Revised online 2016), describes the content of art databases; and VRA Core 4 (Visual Resources Association), describes the content of art database online (VRA, 2007). Additionally, metadata schemata describing primary sources of archival materials, included in this investigation, comprise, ISAD(G) (General International Standard for Archival Description) and EAD V 3 (Encoded Archival Description, version 3).

The primary focus at this point is to investigate the means by which relationships between works/objects in collections may be established, where the related work/object may exist in the same database or in a different database. Next, the investigation of vocabularies uncovers how links can be established to the vocabularies in order to access information stored therein.

Examination of CDWA reveals that it provides three mechanisms to link from a work to a:

a) Related Work Label/Identification
b) Related Visual Documentation
c) Related textual references

Concentrating on a) 'Related Work Label/Identification' denotes "*Works of art or architecture related to the work being described, and a description of the relationship between the works*" The schema provides **text** fields to describe the referenced work and recommends storing all tombstone data of the referred work (preferably system generated). Furthermore, CDWA recommends allocating a specific relationship type and states that the relationships must be reciprocal:

> In record for Sketch A: **Relationship Type:** *study for* **Related Work:** Painting B
> In record for Painting B: **Relationship Type:** *study* [for Painting B] *is* **Related Work:**Sketch A

Available metadata elements facilitate description of the referred object, including elements denoting time span of the relationship, the source of the information about the relationship (a publication or a person) and so forth. CDWA generally expects that both works resided in the same institution, but where that is not the case then the "label may necessarily be free text or the information may be recorded in the DESCRIPTIVE NOTE."

Next, focusing on b) 'Visual Documentation' denotes "*The identification of images that depict the work of art or architecture. Also includes subcategories for an image authority.* " The schema provides approximately 50 elements, **text** fields to describe an image, which may represent a photograph, a digital image, audio or video tape, a slide etc. The content of these elements range from a description of the image, view aspect, date created, technical characteristics of the image (size in kb, resolution, compression), to copyright, to creator and so forth.

Examination of VRA shows that it provides at least two ways to link from a work to a:

a) to any URL
b) to a vocabulary
c) work or a collection or an image

VRA Core 4 provides 9 global attributes that may be added to any element, the result of which depends on the element and attribution combination. The *href* attribute however may be used on its own, as hyperlink to any resource. By way of example, *href* = www.tcd.ie.

Vocabularies may be used to populate metadata fields which then provides access to the referenced entry in the vocabulary and to view the information captured for that item (Hedden, 2010). Establishing a link from VRA to a vocabulary is by means of the VRA metadata element and *refid* and *vocab* elements. *Refid* refers to the unique identifier of the referred item and *vocab* refers the specific vocabulary name. For example, to establish a link to the entry for the artist Jack B Yeats in the Getty ULAN vocabulary, the relevant VRA metadata element to select is AGENT (denoting individuals, groups or corporate bodies) and the *refid* = 500002320 and *vocab*= ULAN. Similarly, to establish a link to AAT, TGN or VIAF (Virtual Internationl Authority File) specify the relevant VRA metadata element, the *refid* and the *vocab* name (Table 4.3).

The RELATION metadata data element uses terms or phrases describing the identity of the related work and the relationship between the work being cataloged and the related work or image. Use this element to relate work records to other work or collection records, or image records to work or collection records. If full relational reciprocity is not explicitly recorded in a local database (e.g. only the part to whole relationship is recorded, and not whole to part), it is recommended that the data exporter add the reciprocal value, based on a controlled set of terms (Table 4.3)

Table 4.3 Metadata elements that enable links to be established between an object and another resource.

Metadata element(s)	Description	Example: Referenced Materials
CDWA		
a) Related Work Label/Identification (p.20)	Describes another work that is directly related to the work being described. One text field contains a brief description and credit line that identifies the related work and distinguishes it from similar works. Ideally, this links to a record for the related work, particularly when both works are housed in the same institution. However, if the related work is external the label may necessarily be free text or the information may be recorded in the DESCRIPTIVE NOTE.	Apollo, Pan, and Putto; painting; Francesco Primaticcio (Italian, 1504-1570); 1559-1560; lost, formerly in Galerie d'Ulysse, Fontainebleau (Ile-de-France, France)
b) Related Visual Documentation (p.36)	Identifies image(s) that depict the work of art or architecture. Metadata elements may describe a photograph, a digital image, a slide, a negative Optionally, Identify the works of art or architecture depicted in the image.	Grandes Chroniques de France, Jean Fouquet; colour slide; 1999; Ansichten und Porträts, Bildarchive Foto Marburg (Marburg, Germany); 00075568,T:0012 e.g. IMAGE TITLE/NAME (or IMAGE VIEW), semi-colon, IMAGE TYPE, semi-colon, VIEW DATE, semi-colon, IMAGE REPOSITORY (with geographic location), semi-colon, IMAGE REPOSITORY NUMBERS.
c) Related Textual References (p.40)	Citations to sources of textual information related to the work being described, including: published bibliographic materials, Web sites, archival documents, unpublished manuscripts, and references to verbal opinions expressed by scholars or subject experts. Also includes subcategories for a citations authority.	Grove Dictionary of Art (online edition). Jane Turner, ed. New York: Macmillan Publishing Ltd., 1999-2002. http://www.groveart.com (3 December 1999). Page: Lowenthal, Anne W., "Claesz, Pieter," accessed 3 March 2005
VRA		
href	hyperlink to another resource e.g. a web page or a file	https://www.nationalgallery.ie/
AGENT	Type	Personal
refid	Reference code or identifier used in the referred resource	500002320 (ID for Jack B Yeats)
vocabid	Name of vocabulary	**ULAN**
href	hyperlink to another ULAN page for Jack B Yeats	http://www.getty.edu/vow/ULANFullDisplay?find=500002320&role=&nation=&prev_page=1&subjectid=500002320
STYLE PERIOD	Type	administrative
refid	Reference code or identifier used in the referred resource	300021147 (ID for Baroque, style-period)
vocabid	Name of vocabulary	**AAT**
href	hyperlink to AAT, Baroque	http://www.getty.edu/vow/AATFullDisplay?find=Baroque&logic=AND&note=&english=N&prev_page=1&subjectid=300021147
LOCATION	Type	administrative
refid	Reference code or identifier used in the referred resource	7001306 Dublin city, Ireland)
vocabid	Name of vocabulary	**TGN**
href	hyperlink to TGN, Dublin city, Ireland	http://www.getty.edu/vow/TGNFullDisplay?find=dublin&place=&nation=&prev_page=1&english=Y&subjectid=7001306
	Type	Personal Names
refid	Reference code or identifier used in the referred resource	49340985 (ID for Jack B Yeats)
vocabid	Name of vocabulary	**VIAF**
href	hyperlink to VIAF web page	https://viaf.org/viaf/49340985/
RELATION	Terms or phrases describing the identity of the related work and the relationship between the work being cataloged and the related work or image. Use this element to relate work records to other work or collection records, or image records to work or collection records. If full relational reciprocity is not explicitly recorded in a local database (e.g. only the part to whole relationship is recorded, and not whole to part), it is recommended that the data exporter add the reciprocal value, based on a controlled set of terms (see table below.)	Personal Names
refid	Reference code or identifier used in the referred resource	500002320 (ID for Jack B Yeats)
vocabid	Name of vocabulary	**VIAF**
href	hyperlink to another resource e.g. a web page or a file	https://viaf.org/viaf/49340985/
ISAD(G)		
Related Materials third element in the 'Allied Materials' section, pp. 1-32	Used to record information about units of description *in the same repository or elsewhere that are related by provenance or other association(s).* Text description, using appropriate introductory wording and explanation of the nature of the relationship . If the related unit of description is a finding aid, use the finding aids element of description (3.4.5) to make the reference to it.	"Sound recordings from the Mabo Papers are held in the National Library's Oral History collection at TRC 3504. (Fonds) National Library of Australia"
EAD		
@href	Hypertext Reference: The locator for a remote resource in a link. When linking to an external file, @href takes the form of a Uniform Resource Identifier (URI). If the value is not in the form of a URI, the locator is assumed to be within the document that contains the linking element.	https://www.nationalgallery.ie/
@linkrole Link Role	A URI that characterizes the nature of the remote resource to which a linking element refers. Data Type: anyURI	<representation href="http://drs.library.yale.edu:8083/fedora/get/beinecke:jonesss/PDF" linkrole="application/pdf">PDF version of finding aid </representation>
IDREF:	ID reference value; must match an existing ID of another element in the document.	For example, the <ptr> element has a @target attribute that can only be an "IDREF," which means it has to reference a valid ID in another element.

4.2.6.6. Observation - Curation of an exhibition

Prior to a pre-scheduled interview with a curator at the National Gallery of Ireland, an opportunity arose to learn first-hand, about aspects of the process of exhibition curation. The gallery was preparing for the re-opening of a wing, after many years of renovation works and they intended to present artworks from their collection that had been in storage for many years. Discussions about the immanent exhibition afforded an opportunity to gather evidence of the curator's ongoing preparations for the upcoming exhibition and demonstrated the curators preferred tools to assist with curatorial tasks of selection and arrangement of paintings. These tools comprise of a hardcopy map of the gallery space, onto which the curator places miniature images of the paintings. The map provides an overview of the complete exhibition space (Figure 4.1, top left), but the curator comments that "*this is more of a reminder of what I plan to do, but of course the actual paintings, they are in your head*."

Since many of the paintings had been in storage for many years the curator informs that currently they are in the process of writing labels and that further research is needed for some of the less well known paintings.

Moving paintings between storage and exhibition spaces (Figure 4.1, top right) requires specialist equipment and foam blocks placed underneath each painting protect the painting at the planned site before hanging the painting (Figure 4.1, bottom right) and several gallery personnel (Figure 4.1, bottom left) are involved in all of these stages. Rearrangement of the painting, in order to reflect the interpretation of the curator and the theme of the exhibition, occurs before hanging, thus minimizing handling and risk of damage. Sharing his professional knowledge, he explains how a curator contextualizes the works through selection and arrangement of the paintings. By way of example, the curator informs that the Vermeer (referring to the painting "Woman writing a letter, with her maid") looks very different when hanging between other works by Vermeer than when hanging with other paintings of women. Alternatively, hanging the Vermeer with other paintings with great light or with paintings from around 1670 changes how we think about and experience a painting and the research shows that curators exercise considerable influence on the visitor's experience at the museum (Moser, 2010; Tröndle et al., 2014; Tzortzi, 2014).

Subsequent to our discussion, retrieving and viewing the images from a documentary (a programme on television), shed further light on the process and highlight the extent of both the physical and intellectual inputs required to curate and exhibition. Contributors to the documentary included Dr. Arthur K Wheelock (Curator of Northern Baroque Painting, National Gallery of Art in Washington, DC, USA) and Dr. Adriaan Waiboer (Head of Collections and Research at the National Gallery of Ireland),

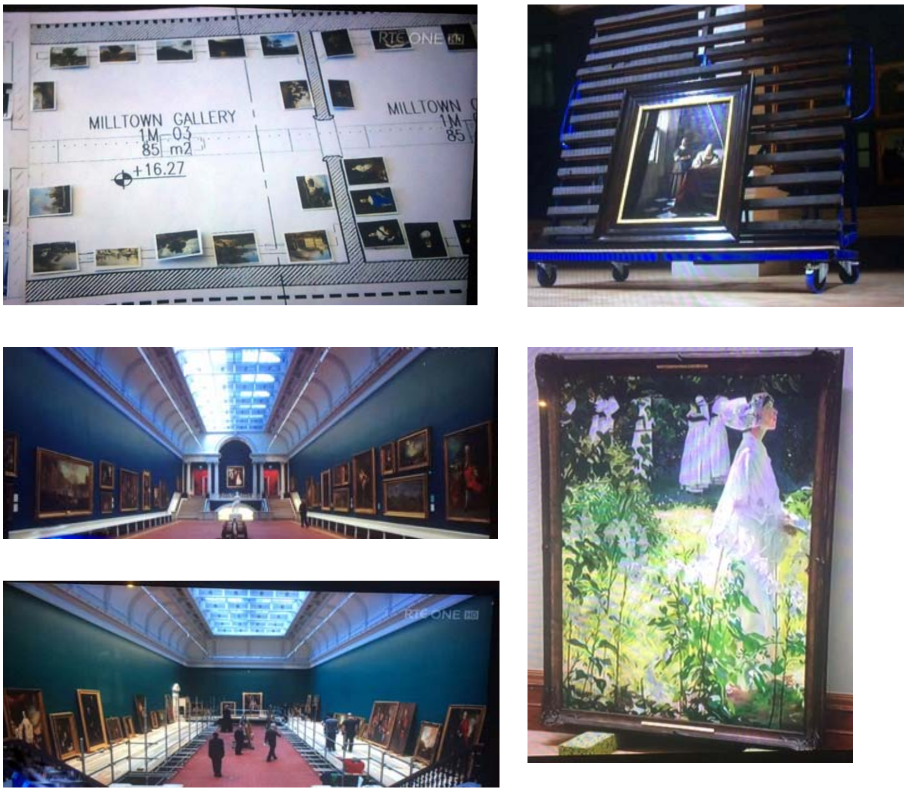


Figure 4.1 Exhibition planning - selecting and arranging paintings for an upcoming exhibition

Top Left - Exhibition map

Top right – moving “Woman Writing a Letter, with her Maid”, Artist: Johannes Vermeer, Dutch, 1632-1675,

Bottom Right - Painting sits on blocks, ready for hanging, A Convent Garden, Brittany, Artist: William John Leech, Irish, 1881-1968

Bottom Left - Preparing for the re-hang, Middle Left -Exhibition space is complete

* All of these images were obtained from a TV documentary about the re-opening of sections of NGI.

4.2.6.7. Museum web page

Viewing web pages on NGI and Tate modern revealed inconsistencies, lack of information and lack of functionality (to read a letter).

1. A review of “The Liffey Swim” on NGI web site shows tombstone data (Title, Artist Name, Year (created), Size, Medium, Provenance and NGI Number) only and a short description (in March 2018) when one search is carried out. https://www.nationalgallery.ie/liffey-swim-jack-b-yeats

2. On the other hand, for the artwork “Men of Destiny” the Title, artist name, artist dates, signed, credit Line, Provenance, exhibition history and label text. This shows that research has been carried out on Men of Destiny.

3. Running a second search immediately on “The Liffey Swim” (from the same area where Men of Destiny was displayed) however presented a different window that included all metadata items shown for “Men of Destiny” http://onlinecollection.nationalgallery.ie/objects/11755/the-liffey-swim?ctx=1d9a10a6-ee50-495b-b511-41ac8fdfab96&idx=0

4. Viewing a letter on Tate Modern Letter from “Paul Nash to Margaret Odeh 25 May 1913”, it was impossible to read the letter and there was no option to zoom up/down on the image and while google chrome increased the window text size, the image size did not increase. http://www.tate.org.uk/art/archive/items/tga-8313-1-1-24/letter-from-paul-nash-to-margaret-odeh/4

5. Viewing “Morning after Rain” on Tate Modern shows the artist name and dates, the title, year and a short biography of the artist (Jack B Yeats). http://www.tate.org.uk/art/artists/jack-butler-yeats-2178
   It did however provide two hyperlinks, one to the Wikipedia page for Jack B Yeats and one to an internal web page that described the style/movement “Expressionism” http://www.tate.org.uk/art/art-terms/e/expressionism

6. Viewing “The Death of Diarmuid, the Last Handful of Water” by Jack B Yeats shows the Artist name and dates, medium, dimensions, Collection, acquisition, reference. It also shows: Inscription, canvas size, purchased details, one exhibition, one literature reference and a description of the work. http://www.tate.org.uk/art/artworks/yeats-the-death-of-diarmuid-the-last-handful-of-water-n05836

4.2.7. Data Analysis Procedures

Data preparation and analysis was carried out as described in (3.7.1.3) and new codes were added as new ideas or themes emerged from the data.

4.2.8. Case Study Findings

The coded interviews were organized and analysed under the following major themes:

- Value of Primary sources of artists archives to art research
- Role of the curator
- Information considered relevant to scholarly research of art
- Resources for art research

4.2.8.1. Contribution of primary sources of artist archives to art research

When asked how archives may contribute to enhancing the intellectual value of an artwork, findings from analysis of the interview transcripts revealed a general consensus that primary sources of artist archives are a valuable source of evidence to the art researcher. The main items contributing to the curators research are itemised (Table 4.4).

Table 4.4 Archival materials contribute different categories of information.

Types of Archival materials	**Object Provenance**	**Artist Methods**	**Context /Artist**
Sketchbooks	✓	✓	✓
Photographs	✓		✓
Register of artworks	✓		✓
Correspondences	✓	✓	✓
Diary entries	✓	✓	✓
Artist Library (including works authored by the artist + collaborations with other artists)		✓	✓
Miscellaneous ephemera			✓

In addition to gaining insight into the working methods of the artist, the experts believed that background details, documents and objects in the archives, as "*pieces of the puzzle*" (Expert A) can also contribute to enhanced understanding of the artist and the artworks. Moreover, experts B and C highlighted that sketches, photographs and diary entries may speak to the provenance of an artwork and possibly for artwork authentication. According to expert B, not only can you not "*do or write art history without archives*" but also that the value of archives can be based on the information contained within the materials or simply by association of the materials with the artist. For example, he noted that the latest theories in archival science propose that archives can '*capture the entire sense of the artist and the world he or she lived in, the artist's interests, the artist's values*' and considers that this contemporary perspective engenders good art criticism and art appreciation. Similarly, Expert C suggested that "*whether its sketchbook, diaries, ephemera, books that the artist owned, photographs, usually whatever archival material is there you can find a use for it… that everything, every piece of an archive will tell you something about the artwork. It all adds up."* In contrast to post enlightenment western art collections expert D noted the lack of artist archives for the collections in her care, but expressed that "*primary sources are fantastic for increasing knowledge of the collection, that anything related to the works helps tell the history*." Checking the archives held at the institution for any related information about the objects was beneficial and if other archives exist with relevant information then "*absolutely, we would want access to them*".

Notwithstanding the positive contribution of archives to art research, the experts identified some problematic areas. Costs of digitizing images and implementing in computer systems concerned expert A whereas accessibility to high quality images, especially for texts such as letters or correspondences was essential, according to expert B. Furthermore, the degree to which archives contribute to art research depends upon the type of research. Although artist archives may deepen understanding of the artist and his artworks, Expert A claimed that archives may be of limited value to particular research, for example, research into conservation or artistic styles. Finally, emphasising the positive aspects of the combination of both primary sources by displaying archival materials next to a work of art, Expert B informs that "*it gives people an understanding of how curators work, of how historians work, that they use this type of material to interpret the artworks*". On the other hand, expert B points out the difficulty of combining archives with artworks stating "*But they're all catalogues of different standards you see, so that's why integration is so difficult*"

In response to the question seeking expert opinions regarding the presentation of archival materials along with a work of art, two of the experts being more familiar with archive collections, offered their views. According to Expert C the intellectual value of archives resides in the 'whole' of the archives and she believed that maintaining archives as a complete unit preserves that value. However, she noted that rather than overwhelming the researcher with the whole collection, that presentation of individual objects along with a note of the collection in which the objects resides, would best assist the researcher. Similarly, Expert B felt it would be unnecessary to present an archival collection when presenting an object from the collection stating "*But in terms of its presentation it could be taken out and shown separately",* concurring also that showing the collection name would be useful.

Finally, in response to the question about whether they consult archives when researching an artwork analysis of the data shows that each of experts interviewed consult the archives during the course of their research. Expert A had just returned from visiting the archives in Holland, but stated "*In May for example I go to the RKD and spent two days there to go through physical stuff, because online, it's just not there. If it was there I wouldn't, I wouldn't go to The Hague*". Planning an upcoming exhibition, expert B intends to search for "*anything that has been digitized from, relating to William Orpen in terms of his archival material*". Checking the in-house archives, expert D searches for "*anything about our texts that will help to tell the history of the object*" whereas, daily consultation of the physical archives at the gallery and monthly consultations of materials at the National Library remain a significant part of expert C's work. Underscoring the significance of archives in relation to provenance, expert C recounted a "*joint research project between the gallery and an institute in Italy, where researchers of archival materials established a direct line of provenance between an artwork held by the local Jesuits, right back to the Mattay family who were the Roman patrons of Caravaggio and who commissioned the work for their private family collection. They established a direct line of provenance back to them and how it got from Italy to Scotland to Ireland and into the Jesuits*."

4.2.8.2. Experts description of their role (curator)

Analysis of the curator's responses when asked to describe the main functions carried out revealed the multifaceted nature of the role of curators (Table 4.5) Findings, however, distinguish two core functions: a) research of the collection and b) exhibition curation. In addition to these core functions the experts carry out a range of secondary or non-core tasks in the course of their work.

With responsibility for the "*care and research*" of the Northern European collection, expert A distinguished two key aspects of the role, "*the collections and exhibitions."* Describing the collections in his care, he related that the Dutch collection of approximately 215 paintings is, "*by far the most important of those four collections*". Evidence of his extensive research on the Dutch collection comprises "*three or four articles on aspects of Vermeer's work*", a Catalogue Raisonné of Gabriel Metsu and contributions to gallery publications. Furthermore, the curator's research extends beyond collections in his care, for example, he stated that "*out of a strong desire to do something different than Dutch, Flemish or German art"* he took the opportunity to research Finnish art, resulting in the publication "Northern Stars and Southern Lights". Additionally, the curator conducts research to build knowledge of individual artworks in the gallery's own collection and for exhibitions at the gallery. Evidence of the expertise and influential nature of the curator's role emerged when he stated "*the collection at the gallery is of such a nature that you can tell the story of Western European Art*", however "it's *the curator who decides which particular story to tell*". Mechanisms for controlling the narrative commence with theme selection, for example, "*the story of Dutch cityscapes"* or "*letter writing"*. Moreover, the influential role of the curator over the audience extends through re-contextualising the artworks. Altering perspectives of artworks, achieved through combining artworks from different collections and or including artworks across a specific time period, a particular nationality or location or some other selection criteria; grouping particular artworks together; the positioning of artworks in relation to each other; and placing a painting centre stage or establishing some focal point.

Table 4.5 Curator tasks

(C denotes a Core task or function, O denotes other or non-core functions carried out from time to time).

Task	**Expert A**	**Expert B**	**Expert C**	**Expert D**
Research of the collection	C		C	C
Exhibition	C	C	C	C
Assist others with Research	O		O	O
Gallery Publications	O	O	O	O
Publish research	O	O	O	O
Attend conference	O	O	O	O
Present Lectures, Workshops	O	O	O	O
Cataloguing			O	

Other tasks of the curator comprise, attending conferences and assisting visiting researchers. A major project was underway at the gallery when this interview took place, with much of the curator's efforts dedicated to this: "*for example, we are now writing labels for the new re-hang in 2017 when we are going to reopen*", requiring further work for lesser known artworks, but he viewed this as a good opportunity for research "*to sort of go back to this work of art, have a good look at it and see what's been done and sort of carry out a little bit more research in order to see if we can find out a little bit more information*."

In his capacity as curator of exhibitions, Expert B outlined the process as: he first selects an artist and a theme for the exhibition and following committee approval, he commences research of archives and digitized materials to support the theme. It emerged from this interview that the gallery curates exhibitions of artworks from other institutions. For example, focusing on his next exhibition, he informs that the artworks reside in external collections "*a lot of the paintings will come from the Imperial War Museum so I'll be using their systems, website*", and that his research will also comprise consultation of archival materials and auction house websites.

Responding to the question of describing her role, in the position as research fellow, art research remains central to the role of expert C, while also "*supporting exhibitions, publications, visiting researchers and other internal projects as they arise*". Her involvement in conducting research for a recent exhibition in collaboration with a contemporary artist highlighted the role of the gallery in developing topical themes of significance to the Irish nation, for example, "*migration ... and the representation of the Irish Diaspora*". Particularly noteworthy was the fact that all of the research for the exhibition used only web based resources. Moreover, the importance for museums to maintain their website content becomes increasingly apparent and the implications for museums further emphasised when she expressed the opinion that new researchers in the field would probably focus more on web based resources and artworks than using printed matter. Regarding the previous observation, it is not surprising that the curator reported that she dedicates part of the time to maintaining the institution's website, as she stated "*We're constantly creating new content for the website. We do a lot more now on the website with exhibitions."*

Describing her role, expert D emphasises "*Care of collections, as an umbrella for taking care of the actual objects and then that involves research, yes, and exhibitions and of course these are all related*." In her opinion, the most significant aspects of the curators job "*is that they ensure that the collection retains its value, that nothing is damaged really... and increases its value through research and exhibitions*". Unlike other curators at her institution, expert D curates exhibitions of objects exclusively from the collections in her care and while the galleries themselves remain the same in terms of their themes and their narrative, the objects of the thesis change. Rotating exhibits not only promotes the collection, especially since "*at any one time only one percent of the collection is ever on display*", but given the nature and fragility of some of the materials, such as paper, papyrus and cuneiform tablets, it also attends to their physical care, by limiting the time of exposure. Writing papers, attending conferences and exhibitions and assisting visiting researchers feature among other tasks carried out along with her research *"I tend to do a bit of research every day, even if it's just writing labels for object rotation"*. Ordinarily, the administration department catalogues the collection, but, tasked with this for an interim period, revealed an enhanced awareness of the need for data accuracy and the benefits of good cataloguing.

Observation at the gallery revealed how curators used their specialist knowledge to purposively select works of art and purposively arrange and contextualise the works of art and collateral texts in such a way that reflects their interpretation of the works and the exhibition theme. These choices demonstrate how the curators exercise considerable influence on the visitor's experience at the museum (Moser, 2010; Tröndle et al., 2014; Tzortzi, 2014).

4.2.8.3. Data considered by domain experts when researching an artwork

During the course of the interviews curators were asked about the kinds of data they consider when researching an artwork. Findings from analysis of the data revealed that curators are more interested in data *about* the works, the factual and contextual data, rather than data *of* or depicted within a work of art, which is of a more subjective or interpretative nature. Curators consider a number of different kinds of data important to describe a work of art (Table 4.6) excluding primary sources of artists archives.

Tombstone data: This factual and relatively static information, tombstone data consists of "accession number, class and (in most cases) work type, title, creator, date made, materials and techniques, dimensions, and credit line" (Zoller & DeMarsh, 2013). While the authors assert that "a free-text credit line field insufficient for finding information related to Gallery donors", experts A and B classified donor details as confidential data. Donor details, storage location, administrative and financial information held in the physical curatorial files, are otherwise unavailable. In addition to the physical files, the institution's collections management system contains records of both tombstone data and confidential data, from which only the non-confidential data is retrieved for display on the institution's web page, for each painting. Accuracy of tombstone data emphasised by expert B and echoed by expert D underpins search and discovery of objects. However, the assumption that tombstone data is always available is contested by expert D, who noted that in some cases, the creation date or year and /or the artist or creator is unknown while in other cases multiple creators may be associated with an object: "*dealing with books you may also have a scribes name, another artist name because there could be a number of people working on the object*."

Table 4.6 The kinds of data (excluding archives) considered by curators when researching a work of art.

(Data types: T = text, D = Date, N = number)

Data	Data Type	Expert A	Expert B	Expert C	Expert D
Tombstone data (Artist, Artist birth/death, Artwork Title, Year Created, Credit Line)	D, T, N	✓	✓	✓	✓
Provenance	D, T, N	✓	✓	✓	✓
Labels	T	✓	✓	✓	✓
Contextual data	T	✓	✓	✓	✓
Exhibition	D, T	✓			
Verso	T	✓			✓
Number of copies	N	✓			
Articles and critiques	T	✓			
Subject keywords	T		✓	✓	✓
Text panels	T				✓
Sitter or people	T				
Christian Iconography	T				✓
Language	T				✓
Translation	T				✓
Style / Movement	N/A				

Provenance: In the art world, provenance is defined as "the history of the ownership of a work of art or an antique (Feigenbaum & Reist, 2012), used as a guide to authenticity or quality; a documented record of this" (OED). Extending this definition, expert A included "*moments when a painting surfaces in a public space*", which is generally at auction but also in an exhibition, possibly on loan from a museum or a private collection. Only one of the experts interviewed (expert A) had conducted provenance research but he maintained that correct attribution can be problematic, that "*provenance of an artwork is not cast in stone*" and that determining authenticity requires research by art historians, where ultimately, the findings reflect "*the researcher's expert opinion*". Highlighting the importance of provenance, expert A discussed the concept further. In the past, the museum (holder of the painting) always determined provenance. However, today, with a shift in the way museums share knowledge, opinions of external experts disputing the provenance of a work, would be considered and accepted by the museum, when proven. Provenance "*reassures the public of the merit of an artwork and legitimises why an artwork is held in the gallery collection"*, according to expert C but aside from the reassurance and justification she expressed feelings that "*knowing the story behind disputed provenance makes the painting even better*". Referring to the collections for which she is responsible, expert D confirmed acceptance of the provenance of the collections, in view of the fact that art specialists and professionals from the British Museum conducted research on acquisitions by the original collector.

Analysis of the data also revealed that all of the curators view auction house websites for evidence of the existence of artworks and to track movements of artworks. The literature underscores that responsibility for provenance research lies with the institution acquiring an object, rather than relying on the dealer or donor to provide provenance documentation (Sandis, 2016). Moreover, implementing strict acquisition policies and ethical acquisition procedures (Levine, 2009; V. Reed, 2013) avoids legal implications and possible compensation or restitution to the rightful owners. Associated concepts raised by expert A, such as ‘number of copies’ and ‘verso’ are discussed in 3.5 below.

Labels: Of primary interest to all of the curators, labels, presented next to an exhibited artwork provide some necessary context to facilitate understanding of the work. Additionally, the curators reported that writing labels requires specialist knowledge and expertise to condense years of research into text on a label. Expert A stated that there “*is a whole debate about labels”*, that over time, labels are getting shorter to increase the probability that people will read the label, whereas “*If you put a label of 500 words you have practically zero people who would actually read it*”. However, they retain all previous versions of labels in their collection management system, for future reference. Working within the adopted standards of between 50 and 60 words, expert D emphasised the need for the text and language “*to strike a balance between being concise enough for a general audience yet rich enough for people who need more*”...”*without alienating either one*”. Although curators are *“never going to be completely uninformed*” expert C felt that curators themselves also appreciate reading labels, especially for unfamiliar artistic traditions or artworks. Describing label writing as “*an exercise in brevity*”, expert D’s institution standard of up to fifty words ensures legibility and limited size, especially given that some objects in the collections, for example, cuneiform tablets, are very small. Also, given that the materials are in different languages Expert D emphasised that some context is needed for increased understanding and concurring with expert C, she stated that she always reads labels by other curators. Furthermore, Expert D highlighted the need for objective description, particularly given her responsibility for religious, Christian collections, which are perhaps less likely understood by other religions and cultures, stating, “*but I try to describe it as an historian, once removed from any sort of belief... because I don’t... want any sort of bias to come through”.* In contrast to the standards for labels for works of art, expert B informs that labels for archival materials tend to be less restricted.

Text Panels: Providing “*a slightly broader overview”* of materials in the collection, expert D stated that they employ a longer text panel, a graphic text panel containing a maximum of one hundred and seventy words, and that “*because I know how long it takes to write text panels, I do read other curators text panels”.*

Exhibitions: Only one of the curators, expert A, voiced an opinion about exhibition metadata for an artwork, stating that it not only allows tracking artworks but also facilitates comparatives in exhibition history, "*it helps to see for example that a painting was part of a very important exhibition... and who lent them, as distinct from provenance*".

Number of Copies: Allied to the concept of provenance, expert A prompted a question about the role of metadata '*number of copies'* of an artwork, when discussing artwork authentication and responded that stating the 'number of copies' ensures that people can distinguish the original painting from the copies. Moreover, he explained, that the likelihood of differentiating originals from copies increases, the longer the duration between the creation of both works.

Verso: Allied to the concept of provenance, *verso,* refers to markings on the back of a painting. Marks; stamps; red wax seals; numbers, an inscription; or monogram may indicate ownership or collection details, according to expert A, and although art historians may not always understand what they mean, "*including this data in artwork documentation is worthwhile as they may mean something to a reader*". Feigenbaum and Reist (2012, pp. 17-18) demonstrate that major collectors frequently mark the back of old-master painting with red wax seals of the owners name and that typically marks are put on the paintings when they enter royal collections. In addition to back markings, other mechanisms for tracking provenance include mounting paintings in custom designed, recognizable frames and placing inventory numbers on the front a painting. (Ibid.) In contrast to back markings on Western art, a Chinese scroll can present a wealth of provenance information (embellished with colophons and seals of the owners) on its front surface. Moreover, these markings are welcome in China, especially from a collector's perspective; they often supplement and enrich the viewer's experience of a painting.(Ibid pp. 32 - 35)

Among other data items considered relevant for research, expert D included broader items than tombstone data such as "*keywords, subject, iconography, cultural affinity- so say something is made in Germany but made by an Italian artist, we'd put that in there"*. Internal information recorded include "*all sort of numbers, so we make sure that acquisition numbers are also in there, if anything is published under a different number... or a name .... So you need to be able to search by that if someone comes in looking for it by that name, say that was published in the 30's by that name or number*." Subject keywords and 'sitter' enhance search and discovery of objects while contextual data (rather than interpretative data) assist understanding of an object according to experts B and D.

4.2.8.4. Resources for research in art

In response to questions relating to the resources that they use for their research, findings from analysis of the data show that curators consult a wide range of both hardcopy, and online resources, in addition to consulting with colleagues and visiting exhibitions (Table 4.7). However, general consensus among the experts concerning all resources is the need for reliable, trustworthy resources and that the researcher should not rely on materials, simply because they are easily accessible.

Each of the curators consult their own collections management systems, paper based files and internal libraries and archives, as expected. Interestingly, especially for this researcher and regarding their internal systems, Expert B reports that the problem when searching for materials is: "*you have to go to different systems to find different things. And because they're so specialised, integration is very difficult. And that's not just a problem here it's a problem in the museum world.*" Continuing this train of thought he references a Discovery Platform: "*where they have one kind of system sitting on top of the other systems that can pull out different things. And yes, it means you get them in the same set of search results but they're still eh, sitting on top, it means you can't go across, if you're here you can't link, can't see what else relates to this.*"

But, besides using their internal systems, analysis shows that curators rely heavily on online resources and Expert A claims that "*The best thing about online resources is that they are searchable*" Considering the wealth of online resource available each of the experts identified particular resources that they continually return to.

Expert A states "*And for Dutch artists in particular I often use a database that is particularly strong on Dutch artists. It's RKD images, RKD artist. So, because, if they find a new burial date for an artist, whatever, it's likely that RKD wants to know, or that the RKD corrects that or that people contact the RKD about it. So the RKD has more up-to-date information*."

According to Expert C, the Getty provides increasingly rich resources and she states "*I use the Getty a lot. I mean they're putting more and more online, it's fantastic, absolutely fantastic resource. They've recently started publishing out of copyright books, books, catalogues, rare books online and that's just a fabulous resource*". Although, currently she uses other resources, "*Well the sources in my own research that I would use the most are newspapers... just kind of general social history, exhibition reviews eh, so I'm working on an essay for an exhibition catalogue that's about royal processions and depictions of royal visits in Ireland, in art. And I've found so much from the newspapers about."*

For Expert D's rich resources can be found in the British Museum and Library website, "*The British Museum is fantastic for prints. The British library have so many, and, of their illuminate manuscripts, the full manuscript up, which is absolutely fantastic when doing research on illuminated manuscripts."*

Auction house websites prove to be a valuable resource for all of the curators, and the reason for this is best explained by Expert B, who relates "*what people often forget is, that the vast majority of art is actually held in private collections, it's not in public collections. And the only way you ever get to see it is when it passes through the auction houses so things like Artnet become really valuable ways of seeing. It's only a very small percentage of artworks that are in* *public collections*. According to Expert C *"Auction house sites, all the time. We always keep an eye on what is coming up at auction. Say if you want to find out other paintings by a particular artist, you get all of their catalogues are online now, you can do a keyword search and they often have archives online."*

Table 4.7 Resources used by domain experts for their research.

(D denotes digital, P denotes paper based and all other non digital resources, including consulting with colleagues and visiting exhibitions and archives)

Resource	Type	Expert A	Expert B	Expert C	Expert D
The Museum System / TMS (CMS)	D	✓	✓	✓	
Adlib (CMS)	D				✓
Museum's own Website	D	✓	✓	✓	
Museum's curatorial files	P	✓	✓	✓	✓
Museum's own library and archives	P	✓	✓	✓	✓
Museum publications / Collections catalogues (including older versions)	P	✓	✓	✓	✓
Museums Archives	D & P	✓	✓	✓	✓
Visit Exhibitions in other museums	P	✓	✓	✓	✓
Consult with colleagues	P	✓	✓	✓	✓
Auction House websites	D	✓	✓	✓	✓
Auction Houses (Adams, Whites)	P				✓
Inter-library loans (Copac.ac.uk)	P	✓			
Artstor	D				X
Other museum websites	D	✓	✓	✓	✓
Open Access Images	D	✓	✓	✓	✓
Art Index – includes journals jstor, artstor	D	✓	✓	✓	✓
Pathé silent movies	D			✓	
Newspapers	P			✓	
Bridgeman	D		✓	✓	
Getty	D	✓	✓	✓	
British Museum	D			✓	✓
The Morgan	D			✓	✓
RKD	D	✓			
Google Books online	D	✓	✓	✓	✓
Art Specialist book and publications	P	✓	✓	✓	✓

Finally, in addition to the many museum websites that the curators consult when researching a work of art, Google Books features as a helpful resource for searching and previewing materials: Expert D *"if I am looking for books to acquire to read, academic books when I am doing research, em because Google books will have partial snippets of them on I can have a little read to see if its relevant"*

4.3. Expert Surveys

This section describes the design, implementation, analysis and findings from a survey of domain experts, involved in the preservation of museum collections and those who carry out art research. The survey questionnaire is based upon questions used in the case study, described above (4.2). Of the 200 emails sent, inviting domain experts to participate in the survey, out of the 60 responses received, 48 valid responses were accepted. This represents a response rate of 24%. Having said that, the number of responses is comparable to existing surveys in the domain, which also inform this study and which are included in discussion of the finding from this survey.

Of the many advantages of web-based surveys (Bryman, 2012, pp. 673-678), specifically for this study, it provided more economical access to a larger population than conducting interviews or a telephone survey. Moreover, it facilitated rapid turnaround in data collection, whereby data were stored directly into a secure database. On the other hand, in the absence of assistance for survey participants, careful questionnaire design and pre-testing, to remove ambiguity and promote consistency of interpretation across the population, was essential.

During survey design the researcher considered the purpose of the survey, whether to use a probability sample or not, the sample frame, the sample size, questionnaire design and response rates (Creswell, 2014, pp. 155-159; Fowler Jr, 2014, pp. 6-11). The sample frame restricts participation in the survey to curators and domain experts responsible for the preservation of collections in their care and to those who carry out research on works of art, thus contributing to the validity of the study.

4.3.1. Purpose

This is the second phase of an exploratory mixed methods research approach and therefore as was the case with the case study, the purpose of the survey is to gain insight into curator's conceptualisation of works of art, how they approach a work of art, and the kinds of things and their relations that curators consider when researching a work of art. In particular, and in light of the increasing interest in artists' archives as potentially valuable resources for art researchers (Breakell, 2015; Koot & Koppenol, 2016; Pyle, 2001; M. Reed, 2017; Sjöholm, 2014) this study explores how and to what extent curators feel that this is the case.

This survey also provides a complementary (quantitative) perspective of the same topics and themes but over a larger population. The advantage of the survey is potentially an enhanced understanding of the domain through data collection from a random and larger population of domain experts; and possibly a more diverse set of experts, by virtue of nationality, culture and/or the range of collections in their care.

Finally, the survey tests the prevalence of findings from the case study over this larger population.

Recall from the case study interviews that seek to understand the domain from the curator's perspective, the survey also focuses on the same interviews topics:

- The role of the curator
- The value of primary sources of artists archives to their art research
- Information about works of art that curators consider important
- The information sources considered relevant to scholarly research of art

Findings from this survey will enhance the researchers understanding of the domain and will assist in refining the research question (Rai, 2017; Yin, 2014, p. 238). Moreover, finding from this mixed methods research approach, will inform the design of a computation representation that incorporates the knowledge of domain experts.

4.3.2. Questionnaire Design

Prior to framing all of the case study questions into a suitable format for the questionnaire the researcher investigated two freely available survey tools and conducted trials based on three of the case study questions. These survey tools; SurveysPro and Google Forms, provided all of the required features except for functionality for timed image overlays. Consequently, the researcher designed a custom survey tool that enabled greater control of text placement, fonts and styles and overlaying of images (on one page rather than extending to multiple pages). The survey program also facilitated direct data capture to a secure, custom designed database, enabling efficient export to .CSV and Excel spreadsheets for statistical analysis.

Next, the researcher selected images of four paintings and eight images of archival materials to include in the questionnaire. Copyright restrictions on the paintings, by Jack B Yeats and the archival materials, displayed on two pages in the questionnaire and also as the background survey image, required a licence. Licenses were purchased for two images of paintings by Jack B Yeats, from IVARO (Irish Visual Artists Right Organisation) Dublin, Ireland, in August 2016 and these were valid for a period of six months only. To avoid extra costs, the researcher swapped out the images of archival materials and selected freely available images from the National Gallery of Art, Washington[40], The Getty Open Content Program[41], and the Rijksmuseum Open Access Policy[42], instead. The relevant questionnaire pages displayed the required statements acknowledging the licences and sources of the open access images.

The questionnaire includes the same questions asked during the case study interviews, albeit presented in a different format, and in a stricter and more logical order than facilitated by the semi-structured interview protocol, used in the case study. The design of the questionnaire incorporates standard practices for questionnaire design. Each question on the survey is optional, as required by the ethics research board at Trinity College Dublin. A textbox,

[40] https://images.nga.gov/en/page/show_home_page.html last viewed 8/8/2016
[41] http://www.getty.edu/about/opencontent.html last viewed 8/8/2016
[42] https://www.rijksmuseum.nl/en/rijksstudio last viewed 8/8/2016

associated with each question, allows participants to provide extra information beyond purely quantitative responses and text is accepted exactly as input by participants with no validation of these entries.

Options identified in the case study are incorporated in questionnaire design. By way of example, if the case study identified ten different types of resources, then the survey includes the ten resources in the appropriate closed questions. Presenting questions and responses in matrices format facilitates rapid response by participants, requiring a single click to select or deselect a response. Single selection of options in matrices restricts input to one qualified option only, by way of example, a curator can select a task as either Core or non-Core, but not both. On the other hand, multiple options can be selected by clicking a qualifier next to the task or resource title as appropriate. Using Likert-type scales (with a maximum of five options) allows participants to (numerically) rate the value of each resource (listed in a matrix) for their research and in some cases to express their agreement with a statement (from strongly disagree to strongly agree) by one click to either select or deselect a response. Text boxes allow entry of free form text of up to 2000 characters and are used for supplementary responses to close questions. A numeric text box allows participants to input the number of years of experience. Finally, combo boxes or dropdown lists (containing lists of countries, education subjects and educational qualifications) standardize responses by allowing data entry by selection, rather than free typing.

Questioning proceeds in a sequential manner, that is, regardless of participant responses, the program presents each subsequent question without branching or bypassing questions. Displaying the page number (for example, Page 1 of 7) at the top of each page indicates the current position in the questionnaire, to participants. Finally, demographics, presented on the last page, use a combination of input methods: click the age range; input a number (for years experience); selecting education level and education from combo boxes / drop down lists of option and typing in their email address and museum name if they so wished.

The survey comprises 7 pages (

Table 4.8) and a Thank You page is displayed when the participant selects to 'Submit' the survey. Of the total 20 questions posed, 9 of these related to demographics and may enable more subtle or nuanced findings. In compliance with the ethics of the Research Board at Trinity College Dublin, the first page presents information about the study and the survey and it specifies the criteria for participation. Likewise, when participants confirm their compliance on the first page, the Informed Consent Form is displayed, outlining the study procedures and participants' rights. Participants are compelled to accept the criteria in order to continue to the survey questions.

Screenshots of the survey are available in Appendix C

Table 4.8 Outline of Survey pages and questions.

Page	Title	Details	No. Of Questions
I	**Welcome: Information Page**	Participants must confirm they have read the information and that they comply with participant requirements	-
ii	**Informed Consent**	Participants must confirm agreement with the terms and rights of participants	-
1	**Role of the Curator**	Core and Non-core tasks	1
2	**Information resources**	Frequency of usage Informative value rating	2
3	**Online resources**	Key characteristics Rate preferences	2
4	**Standards in the domain**	Familiarity with metadata & vocabularies	2
5	**Metadata for Artworks**	3 (superimposed) Images Select important information	1
6	**Considering Archives**	9 images /samples Using archives, Opinion on the value of Archives How and why relevant	3
7	**Demographics**	Demographics (+ email, institution name)	9
	Thank you		

The first question concerns the role of the curator and asks participants to identify the core and non-core tasks involved in their work. The next two questions concern both online and off line resources and asks curators how frequently they use each resource, and to rate each resource in terms of how informative it is for their research. The next two questions concern online resource only, where participants are asked to select characteristics of online resources that are important to them, when conducting their research while the second question allows participants to select and rate the degree to which particular online resources supports their research. Question 6 refers to standards used in the domain (for example, CDWA, Getty ULAN, AAT etc.) and asks participants how familiar they are with the standards and then asks if they are aware of which standards are used in their institution.

Next, participants are asked to identify the kinds of things and the relations they consider important when researching a work of art (Figure 4.2). To assist participants in answering this question, options to examine three different images are provided, that included only the minimum of tombstone data, displayed beside each image. The next set of questions relate to artist archives: and seek curator opinions on the value of archives for their research; how archives might contribute to their understanding of a work of art (Figure 4.3) and the last question asks if and how they have ever consulted archives in the course of their research. The final page seeks demographic information: age, gender, nationality, role, years experience, nationality, education (subject), education level, email address and institution name.

Since each question on the survey enables participants to input free text comments, supplementing responses to the closed questions, these data are analyzed in combination with the quantitative results and in some cases the quantitative results were adjusted to reflect the reality of the situation (full details are provided in each relevant section).

4.3.3. Sample Frame & Sample Size

The sample frame was restricted to domain specialists in museums who carry out research of works of art and those involved in the preservation of aesthetic objects in museum holdings. The sampling procedure was random and participants were unknown to the researcher. The researcher established a random selection of 200 prospective participants by:

- Trawling through museum websites for names and email addresses of curators
- Emailing ICOM (International Council of Museums) who have 35,000 members [43] via the contact page on their website, but did not get a response, to either of the two emails sent
- Locating and exploring a web site for “international network of curators of Dutch and Flemish art” [44] provided a set of names and email addresses of curators.

Each curator name, title, email address and museum was logged in an Excel spreadsheet.

4.3.4. Pre- Survey Release testing

The researcher pre-tested the questionnaire with nine people prior to releasing it to domain experts. The first tester holds a PhD (Photography) from the School of Media, from Dublin Institute of Technology (DIT), Dublin, (2012). The second person holds a Degree in Fine Art from Moscow University, a Masters in Fine Art in the Digital World from NCAD (National College of Art & Design, in Dublin) and a Masters in Computer Science (2016) from Trinity College Dublin. Four PhD students and two Masters students in Computer Science and one person with little computer experience, also took the survey. The testers represented a range of nationalities, comprising, of Irish, English, Russian, Greek and Chinese. The purpose of these trials was to ensure that the survey was suitable for a multicultural population of domain experts and that it was possible to complete it within the indicated time frame. Feedback from testing resulted in a range of amendments: combining, splitting, rewording, reordering and removing some questions and rearrangement of topics in two of the questions. Reformatting the response area to matrices for some questions enabled more efficient data measurement and response analysis. Subsequent to implementing these changes, along with database read/write validation and testing database extraction to Excel spreadsheets, re-testing was arranged. The first two testers confirmed the questionnaire was fit for purpose. Following these tests, template forms in Excel spreadsheets were prepared, in readiness for downloaded responses when the survey was completed.

[43] http://icom.museum/
[44] https://www.codart.nl/

Metadata for Objects in the collection

Title: They Come, They Come (1936)
Artist: Jack B Yeats (Irish 1871 – 1957)
Medium: Oil on canvas
61cm H x 91.5cm W
Signed: Jack B Yeats, lower right
© Estate of Jack B Yeats. DACS London / IVARO Dublin 2016

Click to view other images

Q. 8 Aside from tombstone data to describe an object...

What other information do you consider relevant to your research and knowledge of an artwork?

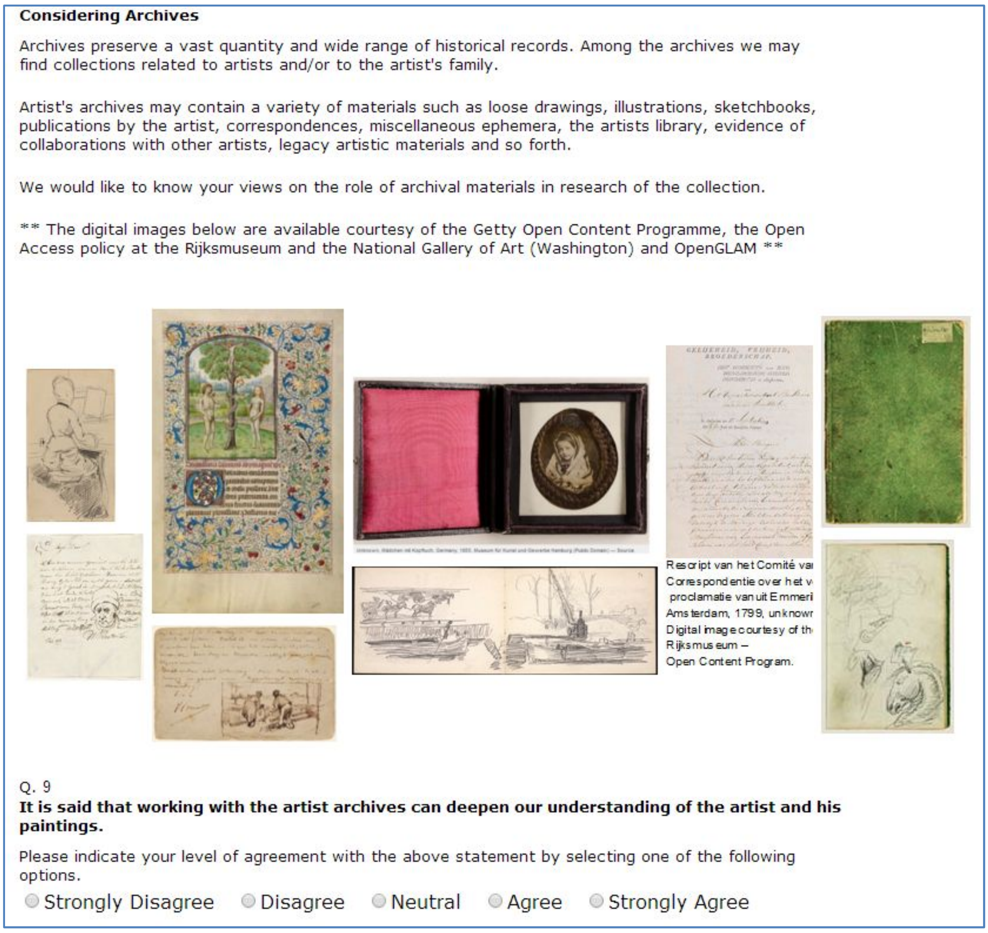
Please select any of the items below by clicking on the text.

- ☐ Description
- ☐ Provenance
- ☐ Current Location
- ☐ Artwork Style / Movement
- ☐ Genre
- ☐ Critical Reviews
- ☐ Publications about the artist / artwork
- ☐ Relationship to other artworks
- ☐ Keywords – theme
- ☐ Keywords – content
- ☐ Keywords – colour
- ☐ No. Of copies
- ☐ Verso
- ☐ Associated archival materials
- ☐ Relationships to other artists
- ☐ Other (Please describe below)

Other: Please describe any other data you believe contributes to your scholarly research and knowledge of the collection.

You may type up to 2000 characters. You can increase or reduce the box size by dragging the

Figure 4.2 The kinds of information considered by domain specialists when researching a work of art.



Considering Archives

Archives preserve a vast quantity and wide range of historical records. Among the archives we may find collections related to artists and/or to the artist's family.

Artist's archives may contain a variety of materials such as loose drawings, illustrations, sketchbooks, publications by the artist, correspondences, miscellaneous ephemera, the artists library, evidence of collaborations with other artists, legacy artistic materials and so forth.

We would like to know your views on the role of archival materials in research of the collection.

** The digital images below are available courtesy of the Getty Open Content Programme, the Open Access policy at the Rijksmuseum and the National Gallery of Art (Washington) and OpenGLAM **

Q. 9

It is said that working with the artist archives can deepen our understanding of the artist and his paintings.

Please indicate your level of agreement with the above statement by selecting one of the following options.

○ Strongly Disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly Agree

Figure 4.3 Domain expert view on the degree to which archives contribute to understanding an artwork

In response to a trial survey invitation email, to two domain experts at the National Gallery of Ireland, the body of the email was changed and the head curator confirmed that their names could be included in the email to the targeted population.

4.3.5. Conducting the Surveys

As there were no changes to the content or intent of the survey arising from the survey testers, the amended survey (Appendix C) was released without the need for re-approval from the research board of Trinity College Dublin. Over the course of the next few weeks individual and personalized emails were sent to the 200 curators, listed in the Excel spreadsheet of participants. Some emails resulted in immediate bounce backs, indicating invalid email addresses and automated responses indicated curators were on extended holidays.

Over course of the next few weeks two curators reported problems with loading the survey and the researcher responded, informing that Javascript must be enabled on the user laptop/PC and that it had been tested in Google Chrome and Firefox browsers. Examination of participant responses shows that at least one of the curators did complete the survey (this was confirmed by the email address provided in her response) but it was not possible to tell if the second curator responded because not all participants provided an (optional) email address. On receipt of a thank you email from the researcher one respondent emailed back “*Two of my colleagues have also responded. I'm generally not good but your survey required so little free text, it made it manageable ! Good luck with it.*”

Due to the slow response rate two reminder emails were also sent. Of the 200 emails sent, 60 responses were received. Checking the database records showed that the completion date/time was not recorded for all participants, which meant that the curator did not ‘Submit’ the response. Consequently, the researcher eliminated responses for these participants from the response dataset, resulting in a total of 48 valid responses.

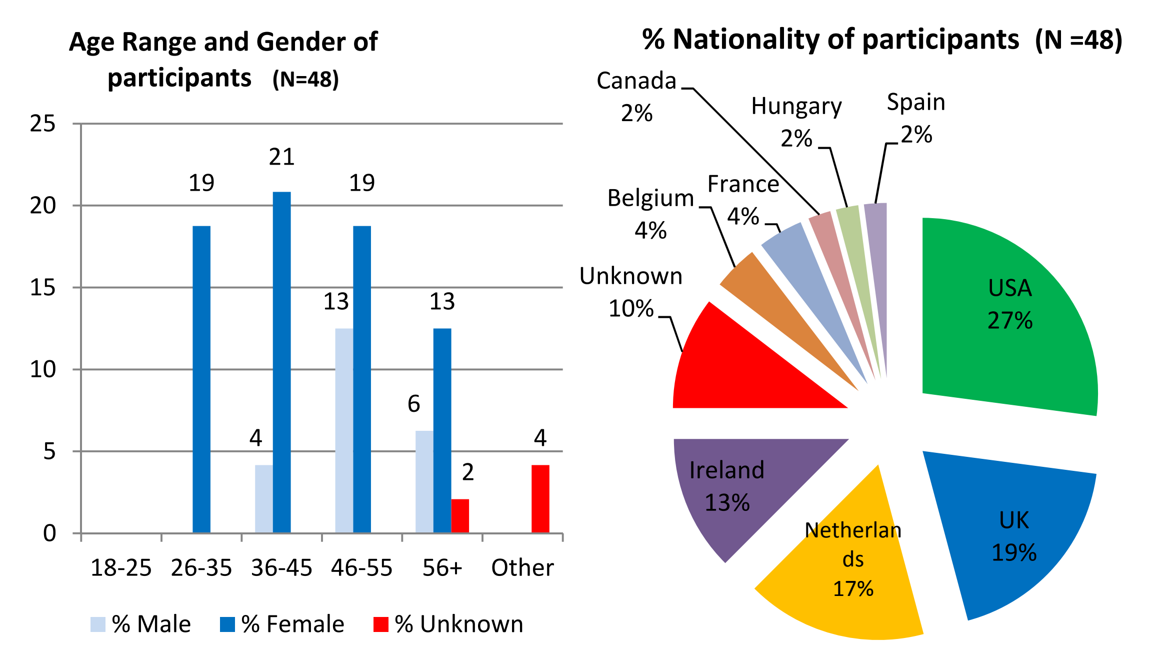
4.3.6. Finding from Analysis of the responses

This section discusses the findings from analysis of the data collected from the questionnaire.

4.3.6.1. Art Experts/Respondents Demographics

Examination of participant demographics, regarding gender and age, (Figure 4.4a) shows that the majority of participants are female. Close to three quarters (69%) are female, whereas less than one quarter (23%) of the participants are male, and 6% are of unknown gender. Moreover, the 26-35 age group comprises females only, while the 46-55 age-group comprises five times more females than males.

It should be noted that the selection of participant for this survey was completely random regarding gender status, as demonstrated by the methods employed for participant selection (4.3.3); also, gender status was not always clear (given the range of nationalities included in the dataset). There are no participants in the 18-25 age-group, the 26-35 age-group is the least represented (19%), the 36-45 age-range comprises 25%, whereas the largest representation comes from the 46-55 age-group, at 31%. The 56+ age-range represents 21% of participants, and a small group, representing 4%, are of unknown age. Taking a cutoff of 65 for the 56+ age-group, participants range from 26 to 65 years, and the average age is 46.15 years.



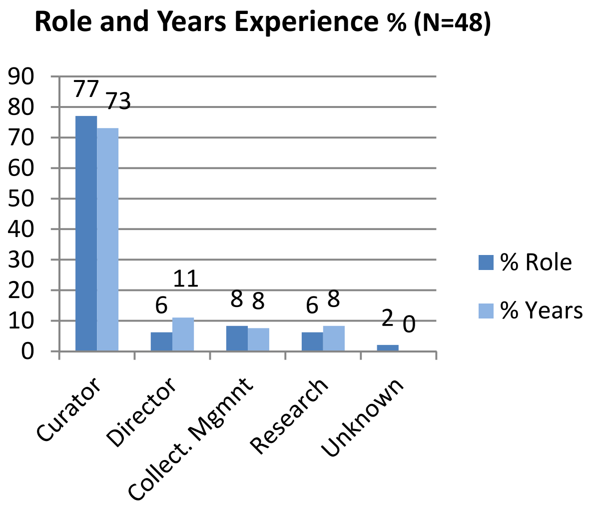
(a) (b)

Figure 4.4 Experts Age and Gender (a) and Nationality (b)

The researcher did, however seek to gain complementary perspectives of the domain through data collection from participants of multiple nationalities, but analysis of the data shows that Russian, Australian or Asian participants did not respond are therefore not represented in this study. Having said that, the survey includes at least nine nationalities (Figure 4.4b) in the cohort of participants, and there may be more, given that 10% are unknown. Even in cases where email address information exists for participants of unknown nationality, the researcher made no attempt to allocate these participants to a country as the evidence from the galleries in Dublin and curator names on websites (by way of example codart.nl) shows that curators, and especially experienced curators, travel the world to work in prestigious museums.

The majority of participants representing just over one quarter of participants are American (at 27%), with the UK the next closest at 19%, followed by the Netherlands at 17% and Ireland at 13%, while Belgium and France represent 4% each and Canada, Hungary and Spain represent 2% each. These figures suggest domination of Western Art traditions in education and approaches to art where, for example, Asian, Australian, Indian and other traditions are excluded from the sample.

To report collectively on the role of the ‘Curator’, the researcher combined selected options ‘Curator’, Curator of Paintings’, ‘Assistant Curator’ and Curator of Research Material and Education’.



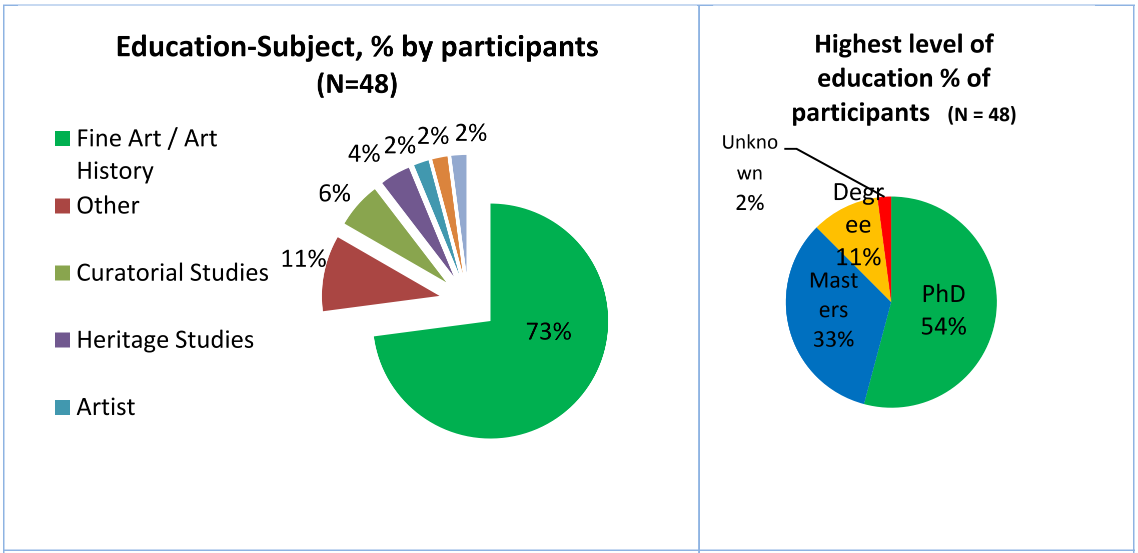
Role	No. of Experts	Total Years Exp.	% Role	% Years
Curator	37	589	77	73
Director	3	89	6	11
Coll Mgmnt	4	61	8	8
Research	3	67	6	8
Unknown	1	0	2	0
Total	**48**	**806**	**100**	**100**

(a) (b)

Figure 4.5 Experts Role and Years Experience (a), Actual Years experience by Role (b)

Analysis of the categories or roles of participants (Figure 4.5a and Figure 4.5b) combined with the age range and years of experience confirms high correspondence between the participants knowledge and the potential understanding and insight to be gained by the researcher, from this survey. With a total of 806 man-years experience in the domain, comprised of: curators, representing 77% of participants have 73% of the experience (589 man-years); directors, representing 6% of participants, have 11% of the experience (89 man-years); and Researchers, also representing 6% of participants, have 8% of the experience (67 man-years).

Equally, analysis of participants' education in terms of the content or subject (Figure 4.6a) and the highest level of education attained (Figure 4.6b), shows that participants possess the formal training for the positions that they hold. Over half of the experts hold a PhD (54%), one third hold a Masters and, 11% have obtained a degree. This, combined with their years experience in the domain enhances the validity of their responses and of this survey. Accordingly, their responses will doubtlessly contribute to the computational representation for this study.



(a) (b)

Figure 4.6 Experts Education by Subject (a) and Highest Level of Educational Qualification (b)

4.3.6.2. The role of the art specialist / curator

Findings from analysis of the data related to the functions and tasks of the art specialist, confirms the multifaceted nature of the role (Figure 4.7) and that the primary task for experts is *research*. 85% of participants regard *research of the collection* as their main task; this is not surprising, especially in light of finding from the case study.

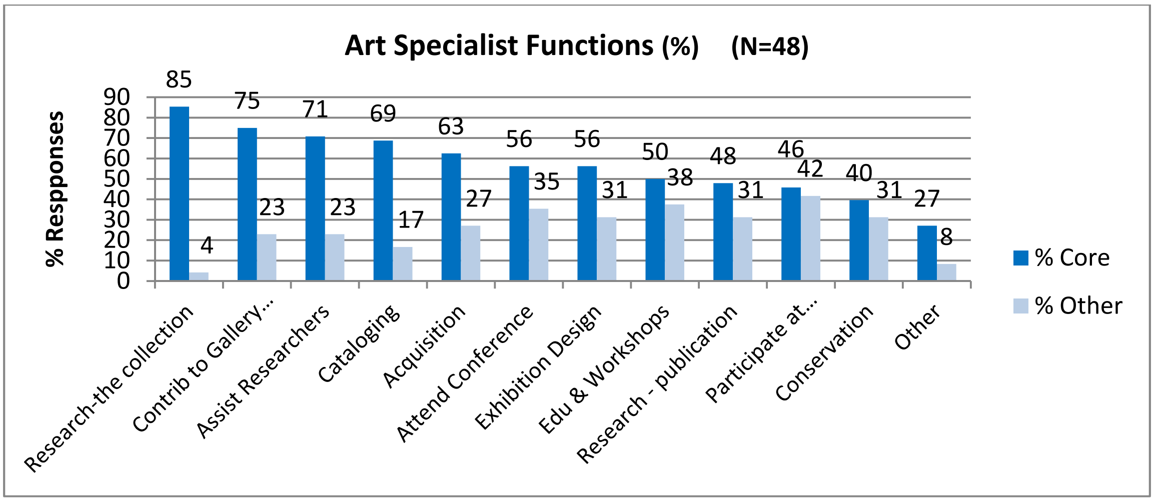


Figure 4.7 Functions of the art specialist, identifying Core and non-Core or secondary tasks

Other research tasks also feature within the Core workload of these domain specialists, comprising: *Contribution to Gallery Publication 75%, Attend Conference 56%, Research Publication 48% and Participate at Conference at 46%.* Checking the details of the 5 participants who do not perform *Research of the collection* shows that they are involved in either *Exhibition* or *Exhibition* and *Gallery publications*. At the other end of the spectrum, a catch-all task entitled *Other* represents a Core task for 27% of participants and a non-core task for 8% of participants, and examination of the comments shows that 4 males and 17 females describe these other tasks, as, for example, organizing loans in and out; relationship management: cultivating donors, liaise with sponsors, lenders and donors; working with artists to organize curatorial projects, liaise with conservations team, work on grant applications and managing curator and research teams. *Cataloging* (69% core, 17% non-core) suggests that these domain specialists are familiar with their institution's collections management systems and may have a keener sense of value or appreciation for metadata and the importance of accuracy and consistency of the data. Surprisingly, *Exhibition* is a core task for just over half of the art specialists (56%) but is a non-core task for another 31%.

Furthermore, checking the responses from the three directors reveals that two (female) directors regard *research of the collection* as a core function, whereas the other director (unknown gender, 40 years experience and a Masters in Fine Art/Art History) regards it now as a non-core task. This draws attention to how participants may interpret questions differently: but based on the 85% already noted above and the findings from the case study, then this would be regarded as a core task.

Moreover, the analysis exercise showed that scrutiny of the comments input by participants is essential. By way of example, focusing on the *Exhibition Design* task, one curator categorized the task as Non-Core, but the participant's comments in the other (free text entry box) functions carried out stated "Curate and coordinate exhibitions" and in this instance the researcher changed the original entry for Exhibition Design from 'Non-Core' to 'Core', to reflect the known reality. However, it must be said that this had no significant impact on the data, and Exhibition Design remains in seventh place within the Core tasks.

4.3.6.3. Information requirements of domain experts

First, it must be noted that the question asking participants about the kinds of information they consider important stated 'apart from Tombstone data' because this is the default or common data shown on many museum websites including such items as Title, Year, Location, Artist, Artist life-span, and Credit Line). Findings from analysis of the data (Figure 4.8) shows that from the 16 categories of information presented, the majority of art specialists consider *provenance* to be the most important piece of information (96%) about objects in the collection. This is not surprising, given that correct attribution assists domain experts in their interpretation and evaluation of the objects, but also because provenance research falls within the realm of the curator's role given the possible legal ramifications (Feigenbaum & Reist, 2012; V. Reed, 2013; Sandis, 2016)

On the other hand, aspects of artworks in terms of style/movement (25%), genre (23%), keywords – colour (8%) and other miscellaneous items (6%) are regarded as the least important, where only a quarter or less than a quarter consider these data when researching a work of art.

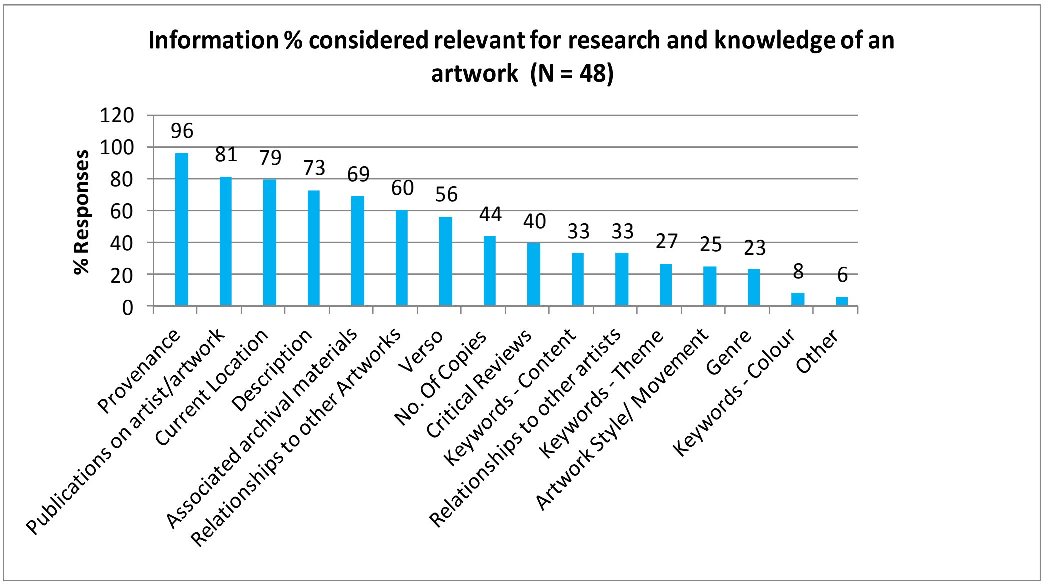


Figure 4.8 Information considered relevant for research and knowledge of a work of art

Publications about the artist and artworks are the next most relevant information for art specialists (81%) and along with critical reviews (40%), shows that art specialists value the writings and opinions of other art specialists as authoritative resources to assist with perhaps alternate perspectives and information that engenders deeper insight into a work of art.

79% of domain experts consider 'Current location' important and this researcher deduces that, that is because it indicates where perhaps the expert may go to either view the actual work of art or which museum (system or curator) to consult for more details about the work or perhaps to organize a loan for a future exhibition. Since the description of an artwork encapsulates a domain expert's interpretation and evaluation of a work of art, it is therefore no surprise that 73% of domain experts regard this as highly. 69% of respondents consider associated archival materials important and this is consistent with the literature showing the potential value of these materials for art research. Furthermore, the relationship to other artworks (considered relevant by 60%) and relationships with other artists (considered relevant by 33%) may provide contextual and comparative information, all of which are important to the practices of art historians and art research (Gombrich, 1995; Panofsky, 1972). Supporting provenance information perhaps may be incorporated in Verso (considered by 56%) and number of copies (44%).

Only one third (33%) of the experts consider 'Keywords-content' as relevant for their research; 27% consider 'Keyword – Theme' relevant; and very few (8%) consider 'Keyword-colour' relevant for their research. In one sense the limited interest in keywords is surprising since keywords facilitate search and retrieval and may uncover relationships based on similar keyword applied to different materials. Perhaps the respondents did not make this connection since the question did not mention 'search and retrieval'. Alternatively, given the possible subjective nature of keywords and especially if they are not used consistently, by employing an authoritative vocabulary such as Getty AAT or TGN, then keywords may not significantly improve search and discovery of a work of art.

A review of the optional comments provided show that a 'photo of verso' is relevant and another expert notes that "*Other museums with the same work (relevant if you work with prints or photographs*". Two art experts consider 'Technical notes & Condition / technical research' important. However, the most important aspect for art specialists researching the collection is the need for up-to date and authoritative sources express by one of the experts as "*Reliability and authority of source*"

In conclusion, the number of information items and the breadth of knowledge about the artworks and their relationships, considered by domain experts, is testament to the complexity of aesthetic objects, from their origination through to the preservation of their intellectual value, by experienced and professionally qualified domain experts. It is also noted that the metadata standards (CDWA and SPECTRUM) facilitate capture of all these information items, albeit, relationships may be established in some instances by textual values only.

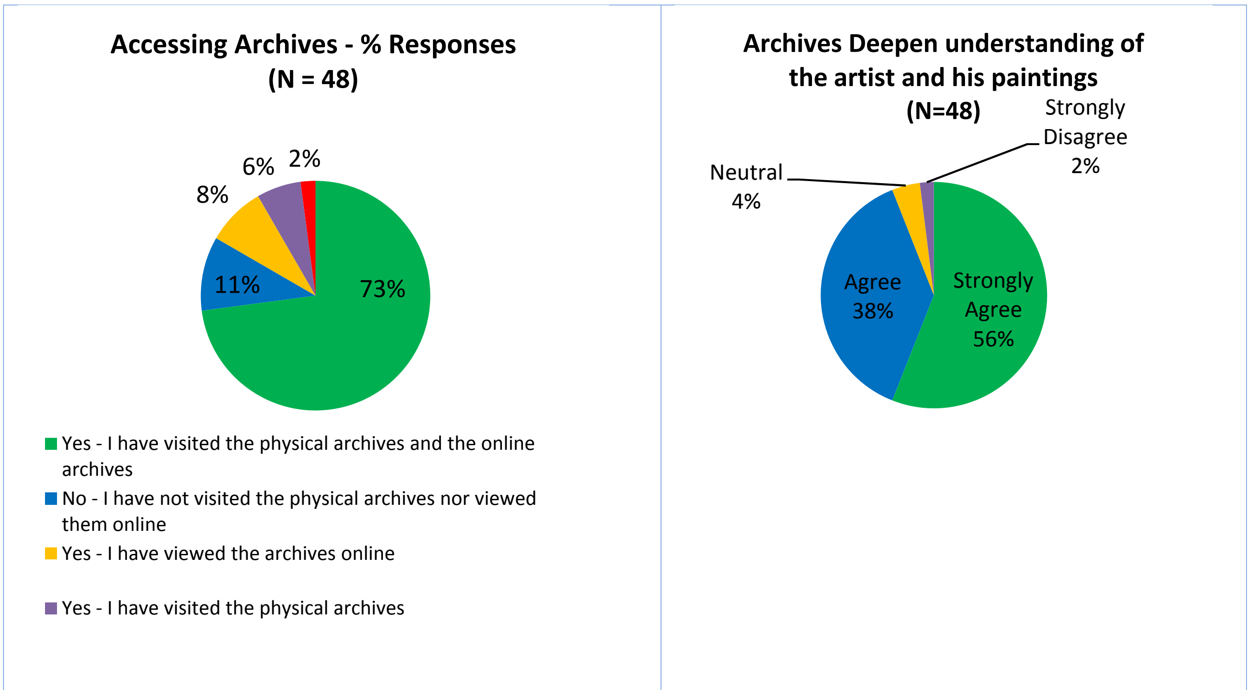
4.3.6.4. The contribution of artists' archives to art research

Results from the previous question in relation to "Information considered relevant for research and knowledge of an artwork", show that 69% of art experts require information from 'Associated Archival materials' when researching a work of art (Figure 4.8). Consequently, one expects that at least 69% of experts recognize the significance of archives for their research.

Analysis of the data shows a strong correspondence between the previous question and these next three associated questions concerning artist archives, but that some anomalies arise also.

Findings from analysis of responses to the question in relation to how the art specialists access archives when researching a work of art (Figure 4.9 a) reveals that almost three quarters (73%) have viewed both online archives and have visited the archives to inspect materials, whereas 11% have not accessed or used archives for their art research. 8% of art experts reported that they have viewed online archives while 6% have visited the archives and 2% (one art expert) declined to respond.

Lacking extra information or comments from the art experts, the researcher deduces that where art experts have not followed up their visit to the archives with online viewing, that the archives are not available online. On the other hand, where experts have viewed online archives, but have not visited the archives suggests that perhaps travelling to the archives is not feasible (due to distance, costs of travel, time or other constraints). However, the responses overall indicate the importance of archives for art research.



(a) (b)

Figure 4.9 Experts access and usage of archives (a) Experts opinion on how archives deepen understanding of the artist and his paintings (b)

Examination of the 11% who have not consulted archives (neither online archives nor physical archives) in conjunction with their responses to Q10 “In what ways do you think that archives can contribute to enhanced intellectual value?”, helps to better understand the reason for this. One expert reported that “Archival sources such as you are referring to and have illustrated do not exist for the material I work with” and another expert noted the importance of archives for art research, but that visiting archives can be an overwhelming experience. Both of these comments appear (framed in a yellow background) in the list of comments from art experts (Table 4.9). Of the other three experts (included in the 11%), one was ‘neutral’ and two ‘agreed’ with the statement that artist archives can deepen understanding of the artist and the paintings, but neither three offered any comment or opinion regarding how archives can enhance intellectual value.

Findings from analysis of domain experts’ opinions on whether artist archives can deepen understanding of the artist and his paintings (Figure 4.9 b), show that a large majority (94%) believe this to be the case: where more than half (56%) strongly agree with the statement and 38% agree with it. These views concur with recent research and implementations on the value of artists archives for art historical research (Breakell, 2015; Koot, 2015; Koot & Koppenol, 2016; Pyle, 2001; M. Reed, 2017; Sjöholm, 2014). However, examination of the data also shows that 4% remain neutral about the value of archives to their research and one expert (2%) strongly disagreed with the statement.

However, attention is drawn to the fact that the initial quantitative analysis showed that 8% of domain experts strongly disagreed with the statement. In addition to selecting the response from a Likert-type scale, a free form text box enabled experts to provide more information or their thoughts on the matter, and 56% of experts input their views. Further scrutiny of these views (Table 4.9), shows that in fact 3 experts who responded 'Strongly disagree' also availed of the opportunity input their views: but their comments contradict that response (refer to the blue shaded comments in the table), and the researcher adjusted the responses to 'Strongly Agree' (at the other extreme of the scale) to reflect the reality of these views.

Table 4.9 Art experts comments about how artist archives can deepen our understanding of the artist and his paintings

The shaded (* blue) comments in the table show responses from art experts who selected "Strongly Disagree" with the statement "It is said that working with the artist archives can deepen our understanding of the artist and his paintings", yet comments from three art experts contradict that selection, and therefore the researcher allocated their responses to the "Strongly Agree" category,. The shaded (** yellow) comments indicate those who have not visited the archives, but who believe that artist's archives can enhance intellectual.

Responses to the question "In what ways do you think that archives can contribute to enhanced intellectual value?"
It isn't only artist archives that are important, understanding when and why a specific object was collected or how it was understood in a particular moment can shed vital light on its biography / changing perceptions of art / the nature of the institution / provenance etc. so a variety of archives have relevance for research.
Provide information that contextualise works of art In the broadest sense possible).
* Archives offer a more direct and often personal view of the artist or issues at hand than general sources can. Sometimes there are connections revealed in archival information that are not available elsewhere.
knowledge of commission; relationship to sitter (portraits); provenance information
** Archival sources such as you are referring to and have illustrated do not exist for the material I work with.
Context of the Art
* The give us access to the unspoken, the detail that is only visible in the intimate, detailed information contained in archives. The finished artworks are key as well, but the archives can allow access to the information behind the artworks.
A work of art can be interpreted on the basis of style, but also on the basis of documents (if available). I have specialized into 16th century artists' archival research not only to flesh out their biographies, but also to help date works, study their creation and iconography,... ideally, pure art historical research and archival research are complementary, but in reality this is rarely the case, and archival research leads a life on its own... see N. Peeters:Family matters: an integrated biography of Pieter Breughel IBelgisch Tijdschrift voor Oudheidkunde en Kunstgeschiedenis LXXVII (2008), p. 45-74., The career of the young Antwerp artist Ambrosius Francken before 1585 Oud Holland 2/3 (2008), p. 99-116.

This table continues onto the next page.

Table 4.9 (Continued) Art expert's comments about how artist archives can deepen our understanding of the artist and his paintings

Responses to the question "In what ways do you think that archives can contribute to enhanced intellectual value?"
In making connections but it may be misleading too in only offering a selection of known related archival material
The personal archives of an artist allows for in depth knowledge on process, research interests, other artists' work as well as surrounding communities.
By being as accessible (inventorized, catalogued, good availability of material, either in digitalized for or physically)
It allows a new context in which to understand the artwork. SOmetimes, it gives insights into materials and techniques. It validates influences, perhaps indicates what an artist was reading, with whom was he/she speaking. Can assist in refining dating of a work. Can help with provenance of this artwork as well as potentially others.
provides more personal perspective of the artist; provides information that has not been published (yet)
Insight into artist inspirations, cultural context, relationships (personal and professional)
Both, altering and/or giving new insights, providing datas.
Archival material gives insight in the prizes paid for artworks, in the provenance of artworks, in the life and network of the artists like no other source can.
Can be influential in altering perspectives on the existing perceived context, can provide a whole host of ideas and questioning.
They can provide a context to the creation of a specific work of art and document the different phases of its production, ones that the artist him/herself may have forgotten.
Authority of readings; accuracy of dating; conformation of attribution, provenance, etc.
New insights, framing the artwork with social and global issues at the time of creation, provide a broader context for the artist's thoughts and choices
It adds another voice and layer of potential information
It is useful to be able to access all the information if it is made available.
Archives provide information about the processes and motivations behind the creation of works
** Researching an artist's archives can be an overwhelming experience, since some artists have an overwhelming amount of material. Although visiting an artist's archive can provide important information, it is best to have an idea what you are looking for before sifting through the material. Archives tend to provide important contextual information about the artist--articles and contemporary opinions about the artwork and/or letters by the artist that provide him/her with a voice about his/her work.
Archives tell the story of an artwork: while interesting and fulfilling, close examination of a painting can only tell you so much, and archival research is needed to tell you where it has been (provenance), about who made it, and the context in which it was created. All of this leads to a deeper engagement with the artwork and its creator, providing new insight in an engaging and informative way.
Understanding of process, technique, subject matter, materials, strategies,ideas and change over time.
Better understanding of artistic practice and ambitions; important data / information

Of the 4% that remain neutral on this question, further exploration shows that one of these experts forms part of the 11% that have not visited archives (previously discussed) and his comments inform that archives such as these, do not exist for the materials in his collection (Table 4.9, the first comment in the yellow shaded background). Lack of comments from the second 'neutral' opinion on how archives can enhance intellectual taken in conjunction with the fact that this expert also forms part of the 11% that have not visited archives (either online or the physical archives) The researcher interprets this to mean that as yet, archives have not contributed value to the experts research.

Further investigation of the one expert (2%) who ‘Strongly disagreed” that archives deepen understanding of the artist and the painting, shows that the expert also did not comment on how archives enhance intellectual value. However, taking into account the data concerning those who have visited archives, examination shows this art expert responded that he/she had visited both online archives and the physical archives. Also, this same expert selected ‘Associated Archival Materials’ as important information when researching a work of art. However, whether the expert selected ‘Strongly Disagree’ by mistake, or visiting the archives contributed little to the expert’s enhanced intellectual value of the artworks, is difficult to judge, especially lacking a comment by the art expert. Therefore, the responses remain as received.

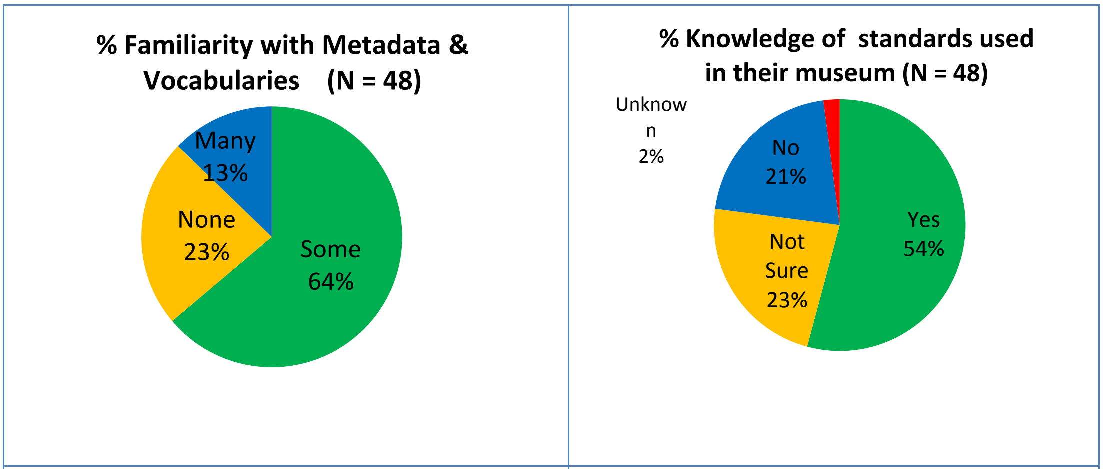
4.3.6.5. Domain experts familiarity with standards developed for the domain

To assist participants in responding to the question with regard to how familiar they are with standards and tools developed for the domain, the survey presents a list of these standards, comprising vocabularies, metadata schemata and Iconclass (iconographic analysis of images).

Despite the fact that work continues on these standards since the 1970’s (vocabularies) and since the 1990’s (metadata standards) results of the analysis of art expert familiarity with the tools (Figure 4.10 a), rather unexpectedly, show that 23% are not familiar with any of these tools.

Of the 64% that report they have ‘some’ familiarity with the standards, one art expert commented that she knows of SPECTRUM (female, 46-55 age group with 10 years experience), one commented that he knows ULAN and Iconclass (male, 36-45 age group with 11 years experience) and a third noted that she was familiar with ICOM and ICN (female, 46-55 with 10 years experience). (Although ICOM and ICN are not standards developed for the domain these organizations promote professional standards and research of cultural heritage). The remainder of the art experts declined to provide optional extra information.

Investigation of the data underlying the 13%, reporting that they know ‘many’ of the standards, shows that it comprises three males and three females. A breakdown of the three males shows that all are in the 45-65 age group: one is from Hungary and has 25 years experience, one is from Canada and has 25 years experience and one is from the USA with 16 years experience. A breakdown of the three females shows that, two are in the 36-45 age group, one from the USA with 15 years experience, one from the UK with 20 years experience and a second female from the UK in the 46-55 age group has 8 years experience. Each of these curators have many years experience, excepting one curator who has eight years experience, and they have been involved in the evaluation or implementation of their institutions’ systems at some point or perhaps more recent professional training incorporates aspects of metatdata or vocabularies in relation to cataloging of cultural heritage objects.



(a) (b)

Figure 4.10 Familiarity with standards and tools in the domain (a), Knowledge of standards and tools used in their museum (b)

Admittedly, participants in this survey comprise curators responsible for preservation of the intellectual value of works of art and are not necessarily responsible for cataloging the works of art. Therefore, their focus is on research of the collection. Having said that, 69% of participants classify Cataloging as a core function (Figure 4.7).

On the other hand, familiarity with these standards equips domain experts with the knowledge that the domain standards facilitate data capture, extending beyond tombstone data associated with objects in the collection. Furthermore, this knowledge may encourage curators to capture additional knowledge in their institutes' systems and be of benefit in the longer term for improved search and retrieval and availability of documentation of the artworks. (Providing their institutes' systems comply with domain standards)

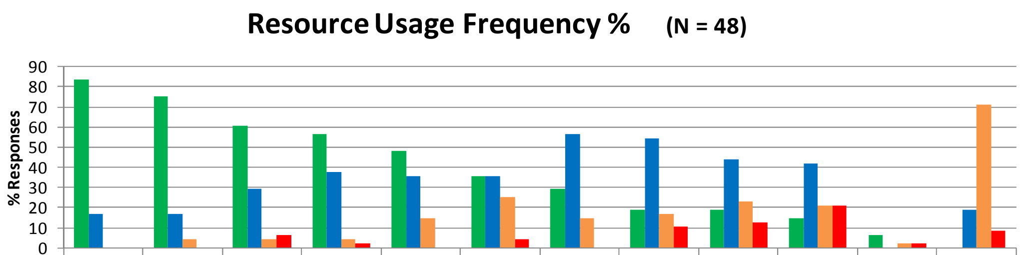
The next question is closely related to the previous one, in that it asks the domain experts if they are aware of the standards employed in their museum systems. Findings from analysis of responses to this question (Figure 4.10 b) show that just over half (54%) of art experts report that they know which standards are used by their systems, while a total of 46% are either not sure (23%) or have no idea (21%) or did not respond (2%). Perhaps these findings are not so surprising, in light of responses to the previous question (Figure 4.10 a). Collections management systems document compliance to standards such as CDWA or SPECTRUM [45], and some of these systems automatically integrate the Getty vocabularies (ULAN, AAT and TGN) but again, this may not be obvious to the art expert, who does not necessarily need to know the particular standards used, provided that their system captures the appropriate information about the collections. However, the results of the survey (Figure 4.12) show that consulting art specialist books and consulting with colleagues rates as more informative than

[45] http://collectionstrust.org.uk/software/ last visited 25/02/2018

their museum systems. Perhaps the systems do not satisfactorily facilitate their information requirements, and they must keep manual filing systems. On the other hand, they may maintain manual systems, and not record the information electronically based on time or other constraints.

4.3.6.6. Resources (manual and online) used by domain experts

Findings from analysis of the data regarding the frequency of use of all types of resources (Figure 4.11) shows that *viewing images online* is the most frequently used resource on a daily basis (used by 83% of the art experts); not altogether surprising given the domain experts surveyed. The second most frequently used resource is the collections management system in the experts' workplace/museum, where 75% of experts use it on a daily basis. The purpose for which they use their museum system is not clear however and one surmises that the experts are possibly reviewing and updating data and or searching for data as part of their research for gallery publications or research for conference or academic publishing. The analysis also shows that some experts use their own institutions hard-copy files on a daily, weekly and monthly basis also.



	Images Online	Inst. Collec. System	Text about images-online	Consult with Colleagues	Art Specialist Books-HCopy	Inst. HCopy files	Websites - other museum	Journal Articles - Online	Art Specialist Books - online	Journal Articles - Hcopy	Other	Visit exhib. at other museum
% Daily	83	75	60	56	48	35	29	19	19	15	6	0
% Weekly	17	17	29	38	35	35	56	54	44	42	0	19
% Monthly	0	4	4	4	15	25	15	17	23	21	2	71
% Rarely	0	0	6	2	0	4	0	10	13	21	2	8

Figure 4.11 Resources used by domain experts and the frequency of usage

Analysis of the data also shows that attending exhibitions at other museums represents a major component of their role (per month), based on the fact that 71% of the experts visit exhibitions on a monthly basis. At least 6% up to 83% of experts consult all of the resources listed, on a daily basis (except and understandably for visiting exhibitions). And it must be noted that offline and hardcopy resources are still very important in the daily work of art specialists: where 56% report that they consult with colleagues, 48% consult art specialist books, 35% consult hardcopy files of their institution and on a daily basis, and 15% consult hardcopy journal articles on a daily basis. The frequency of using these offline and manual resources reduces somewhat on a weekly basis except that hardcopy journal articles usage increase on a weekly basis up to 42% (from 15% on a daily basis. Frequent usage of the manual and offline resources suggests that art experts collaborate with colleagues on a daily and weekly basis and that they set aside time each day, week and month to keep up to date with the latest research published in journals. Free text comments from 5 experts (10%) reported that other resources used comprise, specialist art/history library resources and subscription databases such as Auction House systems, Grove Art online, Getty Provenance Index, RKD online.

4.3.6.7. Rating of the value of resources (manual and online) used by domain experts

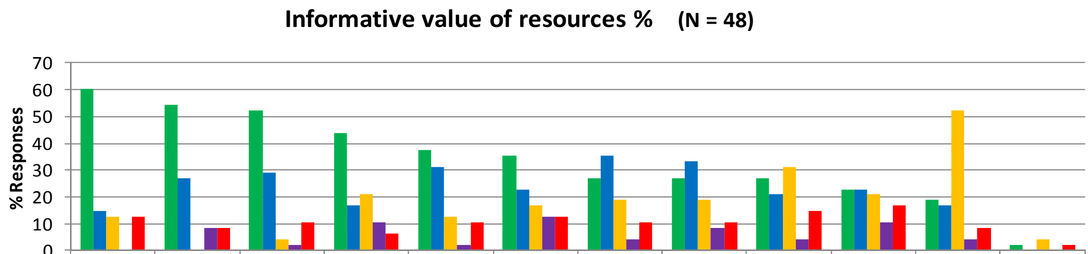
This question asks participants to indicate which of the resources (the same resources used when investigating the frequency of usage in the previous question) they found to be the most informative for their work. Responses utilize a scale measuring the value of the resources, ranging from 5 (the most informative resources) down to level 1 (the least informative). Findings from analysis of the data (Figure 4.12) shows that *images online* represent the most highly informative resource (level 5) for 60% of the experts. Recall also that it is the most frequently used resource (Figure 4.11, at 85 %). For more than half (54%) of the art experts' hardcopy art specialist books, followed closely by consulting with colleagues (52%) comprise the next most informative resources. Next, participants' museum systems feature as the fourth most informative resource according to 44% of the experts. For 38% of the experts, online journals, followed closely by their institutions' hardcopy files, rated as highly informative by 35% of the experts. Equally, resources such as Consulting hardcopy journals, visiting exhibitions at other museums and reading texts about images online, rated highly and as informative, by 27% of the experts. 23% of the experts rate 'Art specialist books online' highly while only 19% rate websites of other museums as highly informative.

Comparison of the decreasing informative value (level 4) with level 5, shows some changes in the figures, where hard copy journal articles move from 27% up to 35%; visiting exhibitions at other museums rises from 27% to 33% and the same number of experts, (as at the level 5 rating) 23%, rate art specialist books online, at level 4 also.

Interestingly, although 19% of participants rate other museums as highly informative (level 5), and 17% consider these as level 4 informative value, 52% of the experts rate websites of other museums with an informative value of level 3, while 4% rate them as not very informative and 8% rate them as the least informative. One observes that these ratings may vary of course, depending on the work being carried out by the domain specialist. But this finding also corresponds with a comment from the research curator from the case study, who noted that much information was missing from museum web pages, as she discovered when carrying out research for an exhibition. The impact of this was that she decided to commence updating her institutions website to include much of the missing information.

Examination of resources classified as low informative value (level 2) shows that only 2% of experts rate consulting with colleagues and journal articles online, at this lower level, and art specialists hard copy books, and visiting exhibitions at other museums are also only considered at this low level by 8% of experts. However, 10% of experts regard their own institution's collections management systems at this low level and 13% feel that their hard copy files are of little informative value, which prompts this researcher to ask, where the information is held or perhaps what kinds of information are required by these art specialists. One explanation for this may be that the expert's research focuses on artworks held in other institutions.

Between 2% and up to 17% of experts felt the need to identify the least informative resources for their work. 17% rated art specialist book online in this category, followed by 15% for images online. 13% classified their institutions hardcopy files in this category whereas 10% of the experts classified consulting with colleagues; journal articles online; hardcopy journals; and visiting exhibiting at other museums as level 1 informative value also.



	Images Online	Art Specialist Books-Hcopy	Consult with Colleagues	Inst. Collec. System	Journal Articles - Online	Inst. HCopy files	Journal Articles - Hcopy	Visit exhib. other museum	Text about images-online	Art Specialist Books - online	Websites - other museum	Other
% Level 5	60	54	52	44	38	35	27	27	27	23	19	2
% Level 4	15	27	29	17	31	23	35	33	21	23	17	0
% Level 3	13	0	4	21	13	17	19	19	31	21	52	4
% Level 2	0	8	2	10	2	13	4	8	4	10	4	0
% Level 1	13	8	10	6	10	13	10	10	15	17	8	2

Figure 4.12 Informative value of resource (Online and Hardcopy) - rated by domain experts

One observes that although all of these resources appear in all categories, from the highest level down to the lowest level (with the exception of images online and art specialist hardcopy books), perhaps the variety of tasks carried out by the domain specialist accounts for this variation of opinions. By way of example, research for an upcoming exhibition may require resources (work of art) outside of the participants' institution; hence, it is unlikely that their institutions files and systems satisfy their information requirements. Similarly, online and hardcopy journals and art specialists books will be of value to those preparing for a conference or submitting an article for publication, but not necessarily required to curate an exhibition.

4.3.6.8. Online Resources used by domain experts

In response to a question 'What do you consider to be the key characteristics of *online* resources that support your role?, the art experts identified and rated the characteristics on a scale of 5 down to 1, where 5 denotes the highest level of support and 1 denotes the lowest value (Figure 4.13). In line with previous responses about resource types used by art experts, more than half of the experts consider *image quality* as the most supportive characteristic (58%) with hyperlinks to other resources as the next most important characteristic for 48% of the experts. A third of the experts rated availability of online resources as key, and a quarter of them identified that description of artworks was necessary. Only 10% focused on the aesthetics of websites and 8% rated functionality such as image zoom and website navigation as key.

Comparison between level 4 and level 5 reveals a major drop from 58% to 13% of art experts that consider image quality important to support their role, at this lower level of support, while hyperlinks and other resources drops from 42% to 25%. The importance of availability of online resources drops to 21%, although the need for online descriptions of works of art increases to 31% (up from 23% considered at level 5) for some art experts and functionality increases to show that 25% of art experts consider that zoom and navigation are important to support their tasks.

Level 3, representing what art experts consider medium level of support, shows a marked increase in the number of participants that consider aesthetics (33%); functionality (33%); description of artworks (33%) as only providing medium support . However, at level 2 (signifying less support for their role) consideration of aesthetic drops to 29% and functionality to 25%.

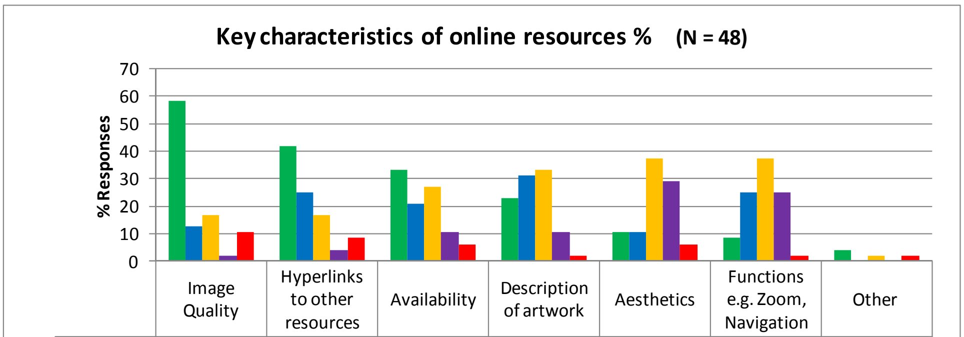


	Image Quality	Hyperlinks to other resources	Availability	Description of artwork	Aesthetics	Functions e.g. Zoom, Navigation	Other
% Level 5	58	42	33	23	10	8	4
% Level 4	13	25	21	31	10	25	0
% Level 3	17	17	27	33	38	38	2
% Level 2	2	4	10	10	29	25	0
% Level 1	10	8	6	2	6	2	2

Figure 4.13 Key characteristics of online resources for domain experts

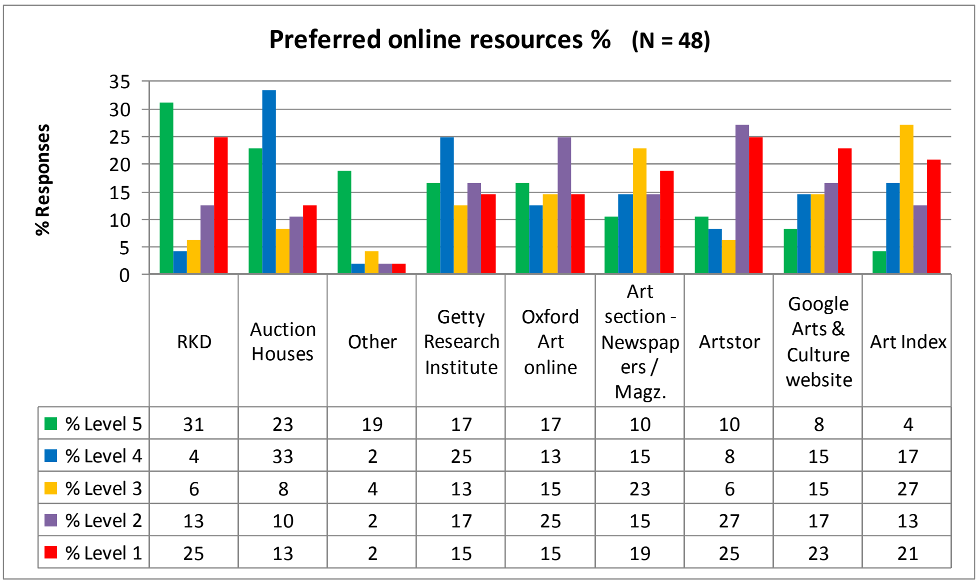
Classified as providing the least level of support to the art experts, level 1, 10% of art experts considered image quality at this level, hyperlinks were considered the least supportive by 8%, and only 6% considered availability of online resources and aesthetics at this lower level. Finally, only 2% of participants considered that description of works of art; functionality and other characteristics should be classified at this lower level.

Analysis of the free text comments provided by four experts provides the data behind the ‘Other’ key characteristic, and includes; the need for higher resolution images (already identified), accuracy of data, ability to download images and data; accessibility; and the quality of possibilities of navigation.

4.3.6.9. Domain experts preference among available online resources

In response to the question “Can you let us know what are your preferred online resources”, by indicating how strongly the resources inform their research process, the art experts identified and rated the resources on a scale of 5 to 1. Level 5 denotes the most favored resources and level 1 denotes the least favored resources (Figure 4.14). Findings for analysis of the data show that there is no premier resource that supports even half of the domain experts. Moreover, data analysis reveals that while some art experts rate one resource as highly supportive of their research process, other art experts rate that same resource as the least supportive. This is not unexpected given the number of participants and variety of nationalities and one assumes a diversity of the collections in their care. By way of example, curators responsible for Western European Art, particularly Dutch art, then RKD most likely, strongly supports their research. However, RKD offers little support for a curator responsible for Islamic art.

Free text comments from 15 art experts provide details of what constitutes the ‘Other’ type of resources (Table 7.1). Analysis of resources classified as level 5 shows that almost one third of the domain experts (31%) identify RKD (The Netherlands Institute for Art History) as the most favored online resource. Auction House website represent the most preferred online resource for 23% of domain experts, followed by ‘Other’ resources at 19%. 17% of art experts prefer both the Getty Research Institute (GRI) and Oxford Art Online, while the art section in newspapers and magazines, and Artstor share the preferred position by 10% of art experts. Google arts and Culture website, preferred by only 8%, is double that of Art Index representing the least preferred resource (4%), at this highest level of classification.



	RKD	Auction Houses	Other	Getty Research Institute	Oxford Art online	Art section - Newspapers / Magz.	Artstor	Google Arts & Culture website	Art Index
% Level 5	31	23	19	17	17	10	10	8	4
% Level 4	4	33	2	25	13	15	8	15	17
% Level 3	6	8	4	13	15	23	6	15	27
% Level 2	13	10	2	17	25	15	27	17	13
% Level 1	25	13	2	15	15	19	25	23	21

Figure 4.14 Domain experts preferred online resources

However, examination of resources classified as level 4 (the next highest category) reveals that one third of the art experts favor Auction House websites, and one quarter favor GRI. 17% of experts favor Art Index, whereas, both the art sections in newspapers and magazines and Google Arts are preferred by 15% of art experts. Preference of Artstor drops to 8% and RKD drops to only 4% (down from 31% at level 5) and 'Other' drops to 2%

Continuing the analysis to Level 3 category (the third highest rated category) shows Art Index in top position, viewed as providing medium support by 27% of art experts. The art sections in newspapers and magazines is viewed as providing medium support by 23% of art experts, and 15% of art experts view Oxford Art online and Google Arts as provide medium support to their research process.

At level 2, 27% of art experts feel that Artstor provides little support for their research process, Oxford Art online is viewed as providing little support, by a quarter of art experts. According to 17% of art experts GRI and Google Arts provide little support. The art sections in newspapers and magazines is regarded by 15% of experts as providing little support, while Art Index and RKD provide little support for 13% of art experts. 10% of art experts identify Auction house websites as providing little support and 2% view Other as providing little support to their art research process.

Analysis of level 1 category shows that a quarter of art experts view that both RKD and Artstor provide the least support for the research process. Google Arts is view by 23% of art experts as providing the least support, while 21% feel the same about Art Index. Almost a fifth (19%) of art experts perceive the art section in newspapers and magazines as providing the least support. GRI and Oxford art online share the least preferred resource as viewed by at 15% of art experts. Auction houses are viewed by 13% experts as providing the least support. Finally 2% of art experts view 'Other' as the least supportive for their research process.

4.4. Requirements Engineering

Before presenting the conclusions of the investigations described in this chapter this section compares the process described above to other approaches in requirements engineering. The discipline of software engineering (SE) has evolved over almost 50 years and the term software engineering was first proposed in 1969 at a NATO conference discussing problems with software development (Sommerville, 2011). Providing a disciplined approach to professional software development, software engineering identifies critical activities to maximize the probability of designing quality software solutions that meet stakeholder requirements in a timely and cost effective manner.

Setting aside all other activities involved in software engineering, according to the Sommerville, the r*equirements engineering* activity represents a crucial aspect in the software development process, the purpose of which is to understand what the system will do and the user expectations of the system. The primary tasks involved in requirements engineering comprise:

a) a feasibility study,
b) requirements elicitation and analysis and

c) requirements specification
d) requirements validation

Moreover, in professional software development the key deliverable from the requirements engineering activity comprises requirements specification document(s) that includes user requirements, system requirements and external requirements in the domain. These documents are fundamental to the future activities and decisions by stakeholders, including the user community, software developers, system architects and the business.

In professional development an initial *feasibility study* would determine the extent to which a system can satisfy the requirements and objectives; the availability of the required technologies and how it may be integrated with existing systems. Whilst a feasibility study was not explicitly documented, the literature review presented details of available technologies, the body of knowledge developed for the domain, and recent implementations and therefore the research proceeded on the implicit understanding that the proposed representation was achievable.

Turning now to the second task, *requirements elicitation and analysis*, Sommerville (2011) defines this as a "*process of deriving the system requirements through observation of existing systems, discussions with potential users and procurers, task analysis, and so on. This may involve the development of one or more system models and prototypes. These help you understand the system to be specified*." Additionally, the author identifies typical methods for deriving system requirements, including, interviews; scenarios; use cases; and ethnography.

Parallels can be drawn between the requirements engineering activity and the processes carried out in this study. The literature review revealed the increasing interest in artist archives for art research and provided the motivation for designing a representation for art specialists and therefore the specific domain; the purpose and the stakeholders were identified at that point. Furthermore, the study was carried out in consultation with art experts at National Cultural Heritage institutions.

To complete this activity, this study employed case study research and survey methods, the rationale for both are described (3.7.1 and 3.7.2). The initial case study research afforded an appropriate approach for this researcher to gain insight into the domain from both the user and domain perspective. Corresponding methods employed in the case study comprised, examination of documentation and standards in the domain; observation of current systems; observations at the gallery (exhibitions represent a core function of the curator's role) in conjunction with several meetings and discussions with expert users, visits to the archives and interviews with art specialists. As the representation in this study largely concerns extending existing systems to incorporate artist archives with artworks, the data (knowledge) represents important aspects for this study and were therefore documented in the case study findings.

From a functional requirements perspective the discussions illuminated the art specialists' views on how best to incorporate artist archives in a system, noting the significance of the context of an archival object, as part of a whole archival collection, that should not be overlooked. Essential requirements also identified included the availability of a web based system with high quality images. Functionality such as Search and retrieval and options to accomplish the requirements were deferred until prototype testing and evaluation.

Sommerville (2011) distinguishes two types of software products: generic products and customised products. Among the generic products the author includes vertical applications designed for a specific purpose or market, such as accounting systems and library information systems. The proposed representation in this study falls within the vertical applications, focusing on art specialists' requirements in the domain as opposed to art specialists in one institution only. Consequently, findings from the case study formed the basis of the survey (described above), to test the prevalence of the case study findings and also to elicit potentially more or alternative requirements from a wider population of art specialists. In one sense, comparison between the survey findings and the case study findings, assisted with '*requirements validation*' in that the evidence from the wider population identified similar knowledge requirements. Furthermore, neither of the findings introduced knowledge requirements beyond those documented in the body of knowledge in the domain.

In professional requirements engineering the next step involves translating the findings from the *requirements elicitation and analysis* into requirements specification document(s). Alternatively, the process may proceed to the development of specific modules or a prototype. This research was carried out solely by the researcher (except where explicitly stated) and therefore rather than generating requirements specifications, the decision was taken to design a prototype. Designing a prototype presents an opportunity to explore how the system works, the extent to which it satisfies the user requirements and amendments required both to the functionality provided and the user interface. Findings from the mixed methods study together with, and informed by further analysis and design of the database schema underpin the design of the prototype.

4.5. Conclusion

The survey results are based on responses from 48 art specialists from nine different countries, representing a total of 806 years experience in the domain. The case study includes interviews with 4 art specialists from 3 different countries with a combined total of 45 years experience.

Findings from the survey corroborate most of the findings from the case study. The case study identified that a core task of the curators involves the preservation of the intellectual value of the work of art and the majority, 85%, of the art specialists surveyed responded that research of the collection was their primary task, while attending conferences and participating at conferences represents core tasks for 56 % and 48 % respectively. However, the case study findings also show that exhibition curation is also a core task of the curator, contrasting with the survey results which shows that only 56% of art specialists report this as a core task (lying in seventh place of the core task identified). Perhaps this variance occurs due to the greater number of curators in some of the larger museums where the respondents are employed and the division of duties is spread over such tasks as contributing to gallery publications, assisting researchers, cataloging, acquisition and attending conferences.

Both the case study and the survey show that images online represent a primary resource for all of the art experts and that external information, beyond the work of art is crucial to their research to building knowledge of works of art. Further similarities between the data collected from both methods are reflected in the fact that art specialists value publications about the artist and artworks, art specialist books and consulting with colleagues as important sources for their research. These findings reflect the findings from both older and more recent studies of the research practices of art experts (Beaudoin, 2005; Brilliant, 1988; Stam, 1997) showing that the image remains central to their research and that art specialist consult a variety or resources including publications, art specialist books and consulting with colleagues to build their knowledge of the artworks.

All of the art experts in both the case study and the survey consult a variety of systems including their own institutions systems and online art specialist sites comprising the Getty Research Institute and subscription based auction house and art index feature among many favoured online resources (which differ depending on the objects or collections for which the research is being carried out) whereas back in 2002, a small study by Rose (2002) showed that only 15% of the 15 art specialists used the internet, possibly due to the dearth of online materials at that time, in comparison to today. Other studies acknowledge the prohibitive time and costs associated with documenting collateral information, but suggest that some of these problems may be alleviated by standardisation of metadata and thesauri (Beaudoin, 2007). Beaudoin (2014) claims that a better understanding of how and why images online are used will help to address the gap in online systems development and design.

Of particular interest to our research, findings from both methods show the strong connection between artist archives and the works of art. The art specialists view these primary sources as essential for gaining insight into the nature of a work of art through the background and contextual information about the artist, the inspiration for the artist's work and the artist methods. 69% of survey respondents regard these primary sources as valuable materials, after provenance (96%), publications (81%) and the location of an artwork and the artwork description. However, when the question focused specifically on archives, 94% of survey respondents (comprised of 56% who strongly agree and 38% who agree) believe that artist archives can deepen understanding of the artist and the paintings. The importance of these materials has been identified in the literature (Breakell, 2015; Pyle, 2001; M. Reed, 2017; Sjöholm, 2014) and studies carried out of art specialist practices and methods emphasise the problem of locating materials online, that materials are missing (have no line presence) and also the need for training in how to use archives (Long & Schonfeld, 2014). The value of these archival materials however is identified by Koot and Koppenol (2016) who claim that ephemeral materials represent rich resources for information for art historical research, particularly for information concerning the relationships between artists, dealers and museums.

The findings from this mixed methods study have provided valuable insight into the domain and the information needs of art experts responsible for the museum collections in their care. Beyond the responses from art specialists and the literature, researcher observations at exhibitions at the gallery revealed similar but not all of the collateral information important to art experts. However, a review of the standards documentation by the researcher, including metadata schemata and vocabularies developed for the domain shows that these standards support the information needs of the art experts. Insights gleaned from this study will inform the design of the representation of the objects and their collaterals and the design of a system that will provide access to the aesthetic objects and the collateral information and equally importantly for the art researcher, functionality to link collateral materials from the archives to the aesthetic objects. The design of the representation is the subject of the next section, Chapter 5.

5. System Design and Evaluation

Findings from the study presented in the previous chapter identify the core tasks and the information needs of art experts responsible for the preservation of aesthetic objects held in collections in museums. *Research, to build knowledge of the collections* in their care, represents the primary task for the majority of the art experts surveyed. Furthermore, the study highlights that *images online* represent the principal, and most frequently used resource, when conducting their research. *Art specialist books* (hard copy) and *consulting with colleagues* comprise the next most informative sources for art experts. The provenance of a work of art represents the prime concern for art experts. Underscoring the complexity of aesthetic objects, the evidence also shows that collateral information such as, *publications* about the artist / artwork, the *current location* and *description* of the artwork and *associated archival materials* comprise aspects of the highest relevance to these experts.

We have shown that relational database technology is well suited to model curator knowledge (Chapter 3). The database schema for our representation organizes these complex objects in structures that describe important aspects of interest to the art experts, including associated archival materials that provide background and contextual information, essential for art experts' knowledge of the collection.

The aim of this research was to incorporate the knowledge of museum curators in a computational representation that may support scholarly research of art. With this in mind, we selected a set of visual art, comprising paintings by, arguably Irelands greatest artist of the 20th century, Jack B Yeats (1871-1957), combined with a set of materials from the Yeats archives, held in the National Gallery of Ireland. The objective was to show how schemata designed for the systematic organization and description of aesthetic objects combined with their associated collateral materials from artist archives, can capture the knowledge of art experts and facilitate a more comprehensive representation of the object and of the art expert's knowledge than is typically available.

Using a relational database, we designed a representation, founded upon long established standards developed specifically for the cultural heritage domain, to organise these complex aesthetic objects and their ontological antecedents, in structures describing aspects of these collections that are important to the art expert. Evidence from the study (Chapter 4) informs our representation of the kinds of information that contribute to art expert knowledge and incorporating the required attributes for the artist, the aesthetic object and contextual information from artist archives contributes to the art expert's insight into aesthetic objects and it is noted that these (archival) documents may also speak to the provenance of a work of art (Table 4.9).

Designing an application comprising a front-end graphical user interface, underpinned by the representation, allowed the art expert to search (based on the captured attributes), retrieve and visualize the aesthetic objects and their related archival materials (for instance, a preparatory sketch, a photograph or a letter). Moreover, the art expert may create relationships between an aesthetic object and these collaterals and input their reflections on the nature of the relationship and share these details with colleagues and art experts using the application.

Evaluation of the application involved researcher experiments and queries to search and visualize the objects and to establish relationships between works of art and collateral materials. Evaluation of the representation of curatorial knowledge and the support for art research was carried out through demonstration and using the system, on a one-to-one basis with 3 domain professionals and noting their opinions and feedback.

Synthesizing the findings from chapter 3 and 4, this chapter describes the theoretical and technical aspects of the system. The first section addresses the context and scope of the representation within the larger domain of cultural heritage (5.1). The next section describes our approach to designing a representation that corresponds with important aspects of reality concerning the knowledge of art experts involved in the preservation of the intellectual value of aesthetic objects in their care (5.2). Technical aspects of the system architecture, comprising a presentation layer, the application/business logic layer and the back-end database services layer, are the subject of the next section (5.3). Details of the data and image acquisition, preparation of materials and population of the database with aesthetic objects, artists and collateral materials from archives are presented (5.3). Next, system experimentation by the researcher and demonstration to art experts and evaluation by art experts, is described (5.4). The chapter concludes with a summary of the findings from this chapter.

5.1. Context and Scope of the representation

The research (Section 2.2) and findings from the study in the previous chapter show that art expert practices and knowledge include and extend beyond the intrinsic qualities of aesthetic objects, to the wider social and cultural context of the creation of the objects. Moreover, recent research and implementations reveal increasing interest in artist archives as potentially rich resources for art historical research.

This evidence re-enforces the fact that a 'good' representation (Winston, 1992) that corresponds to the art expert's knowledge and conceptualization of aesthetic objects, not only represents the aesthetic object but also extends to collateral materials associated with the objects. Therefore, the scope of our study encompasses a representation of objects held in collections in museums, comprising an arbitrary set of fine art paintings and collateral materials from primary sources of artist archives, unlike the typical representation on many museum websites, showing only an image of the work of art and the associated tombstone information.

Art experts classify and annotate these complex aesthetic objects based on a range of metadata that describe both the intrinsic qualities and collateral information. The evidence shows the continual development of a vast body of knowledge within the cultural heritage domain, in relation to the organization, description and management of cultural heritage artefacts, from buildings, to fine arts comprising paintings, sculpture, manuscripts, prints, photographs, object d'art and archival documents and objects held in museum collections. These standards comprise metadata schemata, vocabularies and classification systems to organise, annotate and classify the content art databases (Section 2.3). Examination of the metadata standards revealed that some of these standards provide metadata for a limited description of a work of art (for example, Dublin Core, CDWA Lite) and do not satisfy the requirements of this study. On the other hand, at least two of the standards describe the complete range of information relating to objects held in collections in museums, including collections management procedures from object entry, to acquisition and accession, to cataloging, to inventory control, to loans in/out and exhibition, to assessing damage and loss, to collections care and conservation to de-accession and disposal, to list just some of the 21 collections management procedures, described in SPECTRUM and similar metadata in CDWA.

These standards provided valuable insight into the different classes of objects and the diversity of information for object description and control, across the domain. However, our study is not concerned with collections management per se, and the majority of the data extends beyond the set of elements required to capture the essential aspects of art expert knowledge. But, adopting a *subset of the CDWA metadata standards* to describe the aesthetic objects and EAD (Encoded Archival Description) for the archives promotes consistency and future interoperability of the representation. Moreover, applying these standards shows that our representation for paintings may equally satisfy a representation for prints or drawings or other visual cultural images. Describing manuscripts and sculpture perhaps require additional metadata.

Having selected the metadata standards to incorporate the knowledge of art experts, the next priority concerned identifying the specific information required for the representation. The wide variety of information reported in the survey in Chapter 4 extends beyond the scope of this study, but selecting tombstone data: including "class, work type, title, creator, date made, materials and techniques, dimensions, and credit line" (Zoller & DeMarsh, 2013), both describes an object and facilitates search and retrieval based on a range of data. Necessarily, the representation includes collateral materials from artist archives and these comprise at least the 26 elements from the source documents from which they were harvested (Described in Section 0). The schema design will also include verso, no. of copies and exhibition, as identified in the survey.

Next, we investigated how and to what extent the Getty Vocabularies and classification systems (Section 2.3.3.) may inform our design. The Getty Vocabularies comprise a considerable body of specialist and authoritative knowledge, under continual development, since emerging in the 1970's. Libraries, archives and museums use the vocabularies to catalog, describe and classify works of art and for search and retrieval of cultural heritage materials (Harpring, 2010a). The vocabularies are freely available to browse online, but downloading the database for incorporation into our representation required a licence (Ibid.).

The selected alternative, and following successful trials, showed that storing the URL (Uniform Resource Locator) of the ULAN (Union List of Artist Names) vocabulary entry for the artist and retrieving this as a hyperlink on a webpage in our GUI, automatically displays the relevant ULAN record when the user clicks on the hypertext. Otherwise, we may store the ULAN identifier in the artist/agent table. By way of example, ULAN identifies the artist Jack B Yeats as code 500002320 and as Hedden (2010) suggests, capturing this identifier in our schema would provide access to the referenced entry in the vocabulary and to view the information captured for that item. In addition to providing a short description of the artist, including the artist name, nationality, birth and death details, ULAN lists contributors to the record and bibliographic materials including art specialist dictionaries and online resources. This supplementary bibliographic information will enrich our representation of the artist without the need to input or copy these details in our representation (Ibid.). Similarly, if we wish to acquire authoritative information to describe, for example, the style/movement, technique or supporting medium of a painting we may provide a hyperlink to the appropriate AAT (Art and Architecture Thesaurus) web page. The same mechanism applies to TGN (Thesaurus of Geographic Names) to access additional information about place names, by way of example, where an artist was born, lived or died and where a painting was created or if the painting depicts a particular place. An additional authority file, the VIAF® (Virtual international Authority File)[46] provides information on places and people (among other classes of information) and trials using VIAF provided links to additional information, for example, to the Encyclopedia Britannica biographical record for Jack B Yeats.

Since we intend to include an arbitrary painting, as a contrast to Jack B Yeats paintings for our experiments, we selected Vermeer's painting entitled, "Woman Writing a letter, with her maid". In relation to the Vermeer painting, results from a search of Iconclass based on the term 'handwriting', included Notation 4 'Society, Civilization, Culture' contains an index 49L11, described as 'handwriting, writing as activity' and for this researcher, that appeared to describe the Vermeer painting. A subsequent search of RKD online using the search term "woman writing a letter with her maid" retrieved four paintings, including the Vermeer. Although the painting was classified under notation 4, the art experts at RKD, allocated Iconclass codes 42F51, 'maid' and 46E221, letter, envelope'.[47]

[46] https://viaf.org/ Viewed 10/01/2018
[47] https://rkd.nl/en/explore/images/record?query=woman+writing+a+letter+with+her+maid&start=0 Viewed 10/01/2018

Turning our attention to Jack B Yeats, for whom the ‘horse’ recurs in many of his paintings, a search of Iconclass online using the term ‘horse race’ [48] resulted in 9 codes, found in Notation 4. Discounting one code; ‘46C13141(...)’ denoting ‘horse (with NAME of race or kind)’, the remaining codes started with ‘43C2’ followed by up to 4 digits, many of which could apply to his paintings: from race-track; to horse-racing; to wagon-racing, with horses; and so forth. This was somewhat surprising, given that Iconclass is regarded as more appropriate to iconographic analysis of biblical and mythological subject matter (Hildebrand et al., 2009). Furthermore, widening the search term to ‘horse’ resulted in 300 notations. The potential value of assigning Iconclass codes to Jack B Yeats paintings is unclear however, especially given that a search for ‘horse’ in the description of a work of art would likely retrieve a similar set of paintings (in cases where subject terms have not been catalogued for a painting).

By definition, the Index of Christian Art at Princeton University (Section 2.3) offers no added value to our particular selection of paintings. Beyond the metadata and classification systems considered for our representation, a number of classification systems exist, for particular purposes such as, iconographical themes, fictional characters, names of buildings (Baca et al., 2006) and geographical names, GeoNames, which is freely available for download from the web.

At this point, the scope of our representation extends to the main classes or entity types and the metadata discussed above. The classes of data comprise artist/agent, place, work of art/object, style/Movement, technique, medium, label, exhibition and archives. Tombstone data facilitates search and retrieval of aesthetic objects and artist search and retrieval is currently facilitated by similar data (artist name, nationality, birth and death details) as identified in the ULAN record. Archival search and retrieval is facilitated by some, as not yet decided, metadata elements in the archival record. Furthermore, additional information may be retrieved from the Getty vocabularies through a hypertext link (URL) to the relevant web page.

Systematically organizing these classes and metadata in a well structured schema is the subject of the next section on methods of Logical Modeling.

5.2. A representation of Curator Knowledge

Since we have selected a relational database schema as the basis of our representation of aesthetic objects and collateral materials, held in collections in museums, a suite of established and proven methods, tools and techniques designed by database experts over many years, were available to us. Moreover, this particular representation is founded on established domain standards for organization and description of cultural heritage artefacts and database design principles are adapted as appropriate, to meet the needs of these standards. However, each of these methods and techniques apply not only to the cultural heritage domain, but to the

[48] http://www.iconclass.org/rkd/43C1/?q=horse+race&q_s=1 Horse race

representation of any domain that employs relational database technologies. The design of our representation employed the following methods:

- Conceptual modeling
- Logical modeling
  - Database constraints
  - Functional Dependencies
  - Normalisation
- Database creation (or refinement, if already created during Logical Modeling or subsequent steps)

The following sections discuss the design of the representation, in relation to each of these methods

5.2.1. Conceptual Modeling

Conceptual modeling is a fundamental phase in designing a successful representation and database application. It provides the methodology for representation in terms of the real world objects and their relationships that corresponds with real world requirements (Coronel & Morris, 2016, p. 169; Elmasri & Navathe, 2014, p. 201; Elmasri et al., 1985; Kroenke & Auer, 2012, p. 156; Teorey et al., 1986). In designing the conceptual model for our study we employed tools and techniques of the Entity Relationship (ER) model (Chen, 1976) and the Extended Entity Relationship (EER) model (Teorey et al., 1986) to illustrate an abstraction of the real world data and their relationships, in graphical form.

The theory underpinning the Conceptual Modeling process is the simplification of analysis, based on abstract entities and their inter-relationships (rather than on an overwhelming set of data elements) (Ibid.). This process facilitated top down systematic analysis, where we identified the important entities and their relationships, using constructs of *entity types*, *relationships*, and *attributes* from the ER model and the *superclass/subclass* constructs from the EER model.

The entity types in our representation included both concrete objects and intangible objects or concepts. The concrete objects included aesthetic objects, people, label, and collectible objects held in the archives, and the concepts, included generalizations (superclass) of Work type and base concepts of provenance, exhibition and style/movement, location and events. In addition to paintings, the conceptual model depicts specialized entities of sculpture, manuscript and 'other' kinds of works of art (subclasses of the Work type), because these were observed during our investigations.

The relationships and the degree of the relationships depicted in the Conceptual Model were also informed by data collected over the course of this study. By way of example, the "Masterpieces from the Collection" exhibition featured several paintings by the artist Jack B

Yeats and therefore, in addition to identifying the concept of an exhibition, we also deduce that a one-to-many relationship exists between an artist and works of art. Similarly, each artist or agent may be associated with many different locations, and this represents a many-to-many relationship between these two entity types. The Painting, sculpture, Manuscript and ‘Other’ represent specializations or *is-a* relationship to the Work generalization or Superclass, which will describe common attributes, shared by each of the different kinds of artworks. The Archives will comprise items such as Sketch; Letter; Diary, Photograph, Book and a ‘Misc’ entity type for miscellaneous ephemera, which in this case represent objects that are *part-of* an archival collection or *fonds*. Intuitively the Label entity type represents a one-to-one relationship between a work of art and the label, but on different occasions at the gallery, when artworks were moved from one room to another different labels were used.

Attributes were minimally specified in this Conceptual Model and we assumed a primary key that uniquely identified each particular base entity and a composite primary key for dependent or weak entities such as the Label entity. Resolution of these details and schema design is the core task of Logical Modeling process.

This top down analysis process, resulting in a Conceptual Model (Figure 3.7) highlighted the complexities and the heterogeneous nature of collections held in museum. It also laid the foundation for the next steps in developing our representation.

Before proceeding directly to the ‘Logical Modeling’ step decisions about the scope and context of our representation, within this large and complex domain, needed to be addressed. Moreover, consideration of how the representation can incorporate the knowledge of art experts, and the features and functions that the representation should or potentially could support to assist art experts, required further investigation. These matters are addressed in the following section, discussing the context and scope of our representation.

5.2.2. Logical Modeling – Creating a database schema

The objective of this step was to design a schema that described the database in terms of the structures that represent important aspects of a domain and that incorporate the knowledge of art experts. It involved mapping the constructs from the conceptual model, to a logical model that accurately described the real world entities and their relations, at a detailed level which may be used in a system.

For this research, the universe of discourse is the cultural heritage domain and therefore we employed metadata and content standards, developed specifically for the cultural heritage domain. Nevertheless, similar database design procedures and techniques apply, regardless of the specific domain except that the designer adopts the relevant metadata and content standards for that domain, if they exist, or otherwise develops new metadata schema as appropriate.

Creating a logical model of the database builds on the methodology employed in conceptual modeling (3.2) and is fundamental to successful design of the database schema for the domain. The logical modeling process, examines the abstract constructs represented in the conceptual model – the *entity types*, *relationships* between the entity types, the *attributes* (attributes being minimally defined in this case) and the *generalization/specialisation*, and leads to further refinement in order to create an accurate representation of the domain, at a more detailed level. The process again involves Entity relationship modeling and established and proven mapping procedures and techniques developed by database experts (Chen, 1976; Elmasri & Navathe, 2014, pp. 287-301; Teorey et al., 1986) to map constructs from the conceptual model to a logical model. The mapping process necessarily includes functional dependency (FD) analysis and normalization techniques and results in splitting relations into several smaller relations, ensuring the database schema contains well defined structures. Adhering to established database design principles also reduces potential sources of insert, delete and update anomalies. Based on the requirements for this study, inputs to the mapping process comprise:

- The conceptual model (3.2)
- Findings from the mixed methods study (Chapter 4)
- Metadata Standards: Categories for the Description of Artworks (CDWA) and the Visual Resources Association (VRA) Core 4.0 (Baca & Harpring, 2014 -Revised online 2016; VRA, 2007)
- A content standard: Cataloging Cultural Objects-CCO (Baca et al., 2006)

The traditional mapping procedure comprises the following steps:

1. Map regular entity types
2. Map weak entity types
3. Map binary 1:1 (one-to-one) relationship types
4. Map binary 1:N (one-to-many) relationship types
5. Map N:M (many-to-many) relationship types
6. Map multivalued attributes
7. Map N-ary (N>2) relationship types
8. Map specialization or generalization

Given the complex nature of the entity types in the cultural heritage domain, for this exercise the designer first examined the core entities and relationships and completed all steps of the procedure for each one, as appropriate, before proceeding to the next core entity type. The following sections describe mapping the core entity types for the domain.

5.2.2.1. Map Regular entity types – Agent

Core entity types of interest in the domain comprise Agent, Work, and Archives. The Agent entity type contains records for individual persons, artists, and corporate bodies, including recognized artist studios. Similarly, the Work entity type contains individual records for each work of art. On the other hand, a collection such as archives, representing a number of objects from one source, or shared provenance, and organized in some logical structure, is the subject of (Section 3.6) describing the archives harvesting process.

Starting with the Agent entity type, corresponding to a Personal and Corporate Name Authority (Baca et al., 2006, pp. 279-311) and incorporating ULAN vocabulary, this entity type comprises the museums' records of individuals and corporate bodies associated with the arts. The Agent entity type holds details of artists and all non-fictional persons and organizations related to the artist, such as, noted family members, patrons, mentors and other acquaintances. It may also store details of domain specialists such as prominent academics and art historians along with institutions, museums, repositories, libraries and other organizations associated with the arts and archival materials.

Examination of the items describing the Agent (Table 5.1), informed by the metadata and cataloging standards, from a database design perspective, reveals that storing the data as depicted, would not meet normalization standards.

Closer examination of the Agent entity type also takes into account, recommendations from the content standards (Baca et al., 2006, pp. 78, 282-283) that state:

- "*The creator's names, nationality, life roles, and life dates are critical access points*"
- "*Display Biography, Birth Date, Death Date, Note, and Gender should not be repeatable elements. All other elements should be repeatable.*"

Reviewing the listed elements (Table 5.1), reveals evidence of repeating groups of data or multi-valued attributes, and more non-atomic values, compelling several adjustments to achieve 1NF (first normal form). By way of example, *Name*, *Life roles, related people* and *events* all contain repeating elements and *BirthDate and DeathDate* represent composite elements, containing different types of values: dates and locations. 1NF rules require removal of repeating elements, and reducing composite elements to atomic elements. However, in the cultural heritage (CH) domain artist names are recorded in one attribute, by way of example, agentName may be stored in natural order "Jack B Yeats", (display name) whereas the index to support search and retrieval is generally stored in inverted format "Yeats, Jack B" (Baca et al., 2006, p. 289). Alternatively the supporting index may contain the commonly accepted name for renowned artists, by way of example 'Rembrandt' and not "van Rijn, Rembrandt". A review of the ULAN record for Rembrandt shows two entries for 'van Rijn,Rembrandt', whereas there are 15 with 'Rembrandt van Rijns', but non of the entries are marked as the index. This is understandable for the CH domain especially for easier since search and retrieval by art experts and the general public who use the commonly accepted name. Also of note in the CH domain is the fact that dates are generally recorded as year only and do not include day and month. Given the temporal range of CH, exact dates may not be known and do not adversely affect the data, unlike for example medical systems or finance systems where exact dates may be critical to their operations. On its own, *Place/Location* is ambiguous and without clarification, it cannot be fully functionally dependent on the Agent. The likelihood is that *place* refers to where some event occurred, rather than describing the Agent and at this point, removal from the Agent entity type is the most prudent action to take (2NF).

Table 5.1The Agent entity

Agent	**Example**
AgentID	Unique number allocated by the DB or a code allocated by the cataloguer
Type (Person or Corporate body)	Artist
Name	Jack B Yeats, W Bird (pseudonym– as illustrator, graphic artist)
Biography	Text description
Birth Date	29 August 1871, London, England
Death Date	28 March 1957 (aged 85), Dublin, Ireland
Nationality/Culture/Race	Irish
Life Roles	Artist, Painter, Illustrator / graphic artist, Author
Name Source	Refer to a bibliographic source
Date of Earliest Activity	1888
Death of latest Activity	1956
Place/Location	Place associated with particular events?
Gender (optional)	Male
Related People /Corporate Bodies	John Yeats,-Artist, W B Yeats Poet, other family members, John Millington Synge, Samuel Beckett, Oskar Kokoschka
Relationship Type	Father, Brother, Mentor, Influenced by, Influence on etc.
Events	Yearly visits to Ireland, collaboration with J.M Synge, Authored books and plays. Elected member of RHA in 1916
Note	Any free form text describing the Agent, not covered elsewhere

The *Relationship Type* entity does not describe a particular characteristic of an Agent. Rather, it exists outside of the Agent and defines the different types of relationships between the Agent and other Agents. Also, *Role* describes the Agents role or contribution to some activity as opposed to a characteristic of the Agent. By way of example, Jack B Yeats acted in the role of *artist* (painter) for the painting ‘For the Road’; in the role of *illustrator* for Boys Own Paper, in the role of freelance *writer* for Punch magazine and as *author* of several books and plays. Therefore, as multi-valued descriptors for entity types related to the Agent and not innate attributes of an Agent, removal of these elements from the Agent record satisfies 3NF rules.

Consideration of the *Nationality* item shows that the term 'Irish' describes the Agent’s nationality. However, duplication of this term when describing other Agents increases the probability of inconsistent spelling (update anomalies). Similarly, amending the term from ‘Irish’ to ‘Ireland’ requires amending every Agent record containing the term ‘Irish’, to avoid similar anomalies. Finally, removing the last Agent from the database, containing the term ‘Irish’ would result in losing the term 'Irish'. That being said, the conceptual model already depicts Place as a regular entity, corresponding somewhat to the Geographic Place Authority (Baca et al., 2006, pp. 312-330).

Having removed the repeating groups, functional dependency (FD) analysis identifies a combination of Agent Name and Birth Date as a possible candidate key, in view of the fact that Agent Name on its own may not guarantee uniqueness across all tuples/records in an Agent relation (table). However, given that the Agent may be involved in many different relationships, as described above, it is more efficient to retain the one attribute primary key (PK) designated during the conceptual modeling phase rather than forcing each relationship with the agent, to store multiple attributes of a composite key, as a foreign key to the Agent.

Continuing deliberations on the Agent entity type, the concept of unknown artists requires attention. Arising from our interviews with art experts (Chapter 4) is the fact that an artist or creator of a work of art may be unknown. Suggested workarounds for this problem, rather than accept a work of art with no creator specified (using a NULL value for an artist), include designations that contain the location and or time span, such as *Florentine* or *unknown 16th-century Florentine*, (Baca et al., 2006, p. 280) or perhaps a combination of 'unknown' and culture, for example *unknown German*. However, using generic terms as suggested above means that such agents may create multiple different works of art and the authors recommend that museums devise a standard and consistent approach to address the issue of unknown artists.

Furthermore, the interviews revealed that there may be more than one artist/creator associated with a work of art and that each one may contribute in a different role. Therefore, rather than establishing a relationship directly between a work of art and the agent entity, an intermediary relation is needed, such as an 'Artist-role' entity.

At this stage, the process results in nine entity types, corresponding to nine relations (tables) in the database schema. These comprise: Agent, RoleType, Place, RelationshipType, and EventType representing regular entity types, whereas, entity types, AgentLnkName, AgentLnkRole, AgentLnkAgent and AgentLnkEvent represent multi-valued weak entities (Table 5.2). Derived entity types based on the repeating groups of Names, Roles, Relationships and Events represent *weak* entity types, whose existence depends on the presence of an owner or parent entity type. Moreover, the weak entity type always contains a foreign key (FK) equivalent to the parent primary key (PK) of the Agent, satisfying referential integrity constraints. By way of example, removing the one and only painting created by Gabreil Metsu, from the collection and removing the artist Gabriel Metsu from the Agent relation, compels removal of the roles, events, relationships and names associated with Gabriel Metsu.

A closer look at each weak entity is necessary before deciding on the PK, the degree of the relationship between dependents and owners, and if their relationships are mandatory or optional.

- Since an Agent has at least one name (or allocated a name as discussed above), but may have multiple names, this represents a mandatory 1 to N (many) relationship between the Agent and the AgentLnkName *weak* entity.
- Similarly, an Agent participates in at least one event, either the birth of the artist or creation of a corporate body, and therefore the Agent has a 1 to N relationship with the AgentLnkEvent weak entity.
- An Agent may or may not be associated with one or more Agents, therefore the relationship between the Agent and the AgentLnkAgent represents an optional 1 to N relationship.
- The role element describes the life role and/or the contribution to the creation of a work of art, regarding an artist Agent. For other individuals lacking a defined role, the system may allocate a default role. For a corporate body the role denotes the function of the Agent. Therefore, the relation between Agent and AgentRole represents a 1 to N relationship.

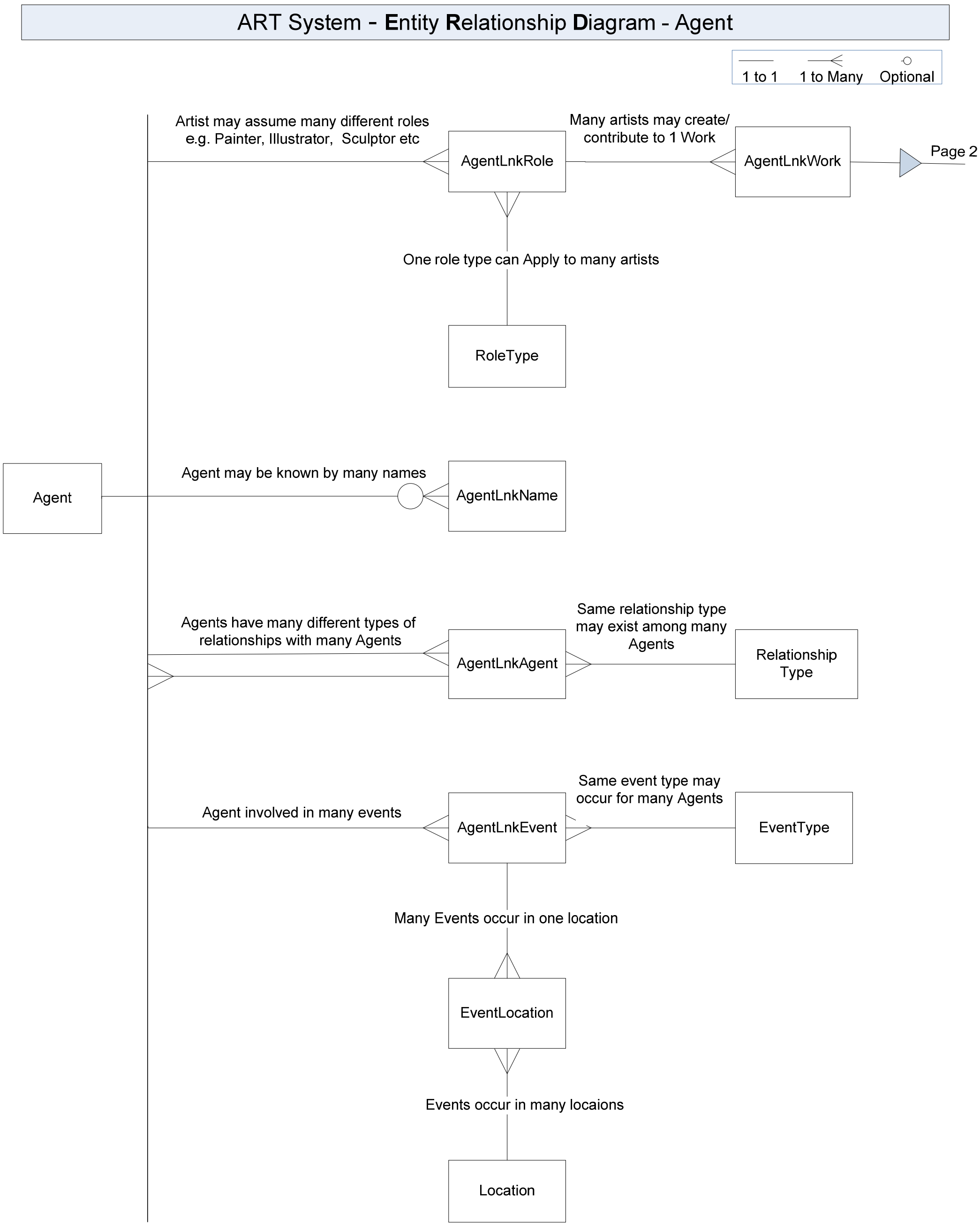


Figure 5.1 Entity Relationship Diagram (ERD) for Agent/Artist and associated relations.

Table 5.2 Database schema – result of logical modeling of Agent

Entity Type	Primary Key	Attributes->									
Agent	AgentID	DisplayName	BirthDate	DeathDate	EarliestDate	LatestDate	NationID	Gender	Biography	Note	Index-Name
AgentType	AgentTypeID	AgentTypeName	Vocab	Refid							
RoleType	RoleID	RoleName	Vocab	Refid							
PlaceLocation	PlaceID	PlaceName	Vocab	Refid							
RelationshipType	reltypeID	relTypeName	Vocab	Refid							
EventType	EventTypeID	EventName	Vocab	Refid							
* AgentLnkName	AgentLnkNameID	AgentID	AgentName	PrefferedYN	Lang						
* AgentLnkRole	AgentLnkRoleID	AgentID	RoleID	roleNote							
* AgentLnkAgent	AgentLnkAgentID	Agent1ID	Agent2ID	reltypeID	relNote						
* AgentLnkEvent	AgentLnkEventID	AgentID	EventTypeID	EventDate	EventNote						

Table 5.3 Database Schema - result of logical modeling of ArtWork/Object

Entity Type	Primary Key	Attributes->											
Object	ObjectID	ObjectTypeID	ClassID	Title	Creation Date	EarliestDate	LatestDate	Height	Width	MeasType	StyleMovID	MaTechID	Inscription
ObjectType	ObjectTypeID	ObjectTypeName	Vocab	Refid									
ObjectClass	ClassID	ObjectClassName	Description										
StyleMov	StyleMovID	StyleMovName	Vocab	Refid									
PlaceLocation	Already Identified in Agent process												
Subject	SubjectID	SubjectTerm	Vocab	Refid									
* ObjectLnkAgentRole	ObjectID, AgentLnkRoleID												
* ObjectLnkSubject	ObjectID, SubjectID												
Genre	GenreID	Vocab	RefID										
Painting	ObjectID	Verso	NoCopies	GenreID									
* ObjectLnkLabel	ObjectID, DateUpdated	Description											

5.2.2.2. Map Specialisation-Generalisation entities – Artwork entity

The content standards documentation describes cataloging of works of art as singular objects and as works consisting of many parts, where a collection and or a series may comprise items such as, installation art, paintings, sculpture and prints and possibly require organization in some logical order. (Baca et al., 2006, pp. 13-19). Therefore, distinguishing between a Work record, a Collection record and an Image record is by means of a record type indicator (Ibid.).

For the purposes of this study, the Work entity type corresponds to a singular aesthetic object, such as a painting, a sculpture, a print or a photograph. Representing materials logically organized in a hierarchical collection comprising fonds, series, file and as individual items, is the subject of the next section, involving harvesting archives from texts and storing them in a database.

Focusing on a singular object, the Work entity type may be described by a number of attributes (based on the metadata and content standards) typically (Table 5.4) shown next to an image of the painting when displayed on a screen and perhaps in hardcopy print.

Table 5.4 Items describing a Work of art. The example shows a painting 'For the Road' by Jack B Yeats

Work / Object	Example
ObjetID	Unique number allocated by the DB
Class	Paintings-European art
Type	Painting (other options are: drawing, sculpture, engraving etc.)
Title	For the Road
Creator	Jack B Yeats (Irish, 1871-1957)
Role(s)	Painter (artist)
Creation Date	1957
Earliest Date	1957
Latest Date	1957
Measurements	61 cm X 91cm
Materials and Technique	Oil on panel
Style	Expressionist
Description	Free text
Description Source	Pyle, H Jack B. Yeats, A Catalogue Raisonné of the Oil Paintings 1992
Location	National Gallery of Ireland
Inscription	Jack B Yeats, lower right
Related Work	a similar tunnel of trees in 'Excitement on a Road' (Pyle, H, 1992, Catalogue Raisonne No. 796, 534, 1085) Registry in archives Page 10
Subject	White horse, man, journey, allegory, purpose

Unlike the Agent entity type which was mapped to a single relation, the conceptual model depicts the *painting* entity type as a subclass of the generalization / superclass *Artwork*. In this situation a number of design options are available:

1) Create a relation for the superclass /generalization *Artwork,* and create an individual relation for each of the specialization /sub class entity types: *Painting*, *Sculpture*,

Manuscript.... *Other*. The same primary key is allocated to the superclass and none, one or more than one of the sub class relation (representing a *total* or *partial* and *disjoint* or *overlapping* sub class). The superclass relation comprises all common attributes and each sub class relation contains only those attributes specific to the sub class or kind of Work that it represents. The primary key in the sub class relation is the same as the primary key in the superclass relation.

2) Create one (disjoint) relation for each sub class relation, containing all common attributes and those specific to the sub class / specific kind of artwork. (This dispenses with the superclass)

3) Create one relation only, containing all common attributes and all specific attributes for each and every type of Work, using a discriminator attribute to denote the kind of Work (or sub class) in each tuple. However, this would probably result in many NULL valued attributes. By way of example, a painting may require attributes for genre, sitter, style/movement, verso, and colour, while a sculpture may require extra attributes for weight and circumference that are not required for a painting.

4) Create one relation only, with multiple type attributes used to denote the Work type and all attributes (both common and specific) that may or may not apply to particular types of Works. This would work for overlapping sub classes and for disjoint sub classes.

For this study, the researcher decided to select option 1, create a superclass containing common elements to all works of art, and an individual relation (table) for each type of work of art. Selecting this option means that relationships can be defined on the superclass and automatically inherited by each specific art kind. Even though this system focuses on *paintings*, selecting option 1 effectively plans for future additional art kinds.

Examination of the items describing a work of art through the lens of a database designer reveals that storing the data as depicted in the ArtWork table (Table 14) does not meet normalization standards. Similar to the decomposition of the Agent entity type, assume for now that a WorkID exists as a unique identifier or Primary Key (PK) for an Object, given that the *Title* may not guarantee uniqueness and a composite PK would otherwise be required. Achieving 1NF (first normal form) standards requires several amendments. By way of example, *subject* contains repeating or multi-valued elements, *measurements* contain both values and measurement type and *Creator* is also a composite field of Name, Nationality, birth and death dates (1NF). Similarly, the object/painting may be associated with multiple objects and with multiple subjects (1NF). *Role* describes the contribution of the creator and does not directly describe the work of art (2NF). Also, although Expressionist may describe the style of the painting, duplication of that term to describe other paintings increases the probability of inconsistent spelling (update anomalies). Similarly, duplication of the term means that if the term ‘Expressionist’ is changed to ‘Expressionism’ then every artwork of that same style requires updating to avoid anomalies. Finally, removing the last ‘Expressionist’ painting from the database causes the loss of the style ‘Expressionist’. That being said, the conceptual model already depicts StyleMov as a regular entity and therefore a foreign key StyleMovID will exist in the Object relation/table.

Findings from the mixed methods study showed that some curators are interested in the ‘no of copies’ of a work of art, the verso of a painting and they sometimes wish to keep a record of the different version of labels (captions) that they write for a work of art. The No. of Copies and Verso attributes are assigned to the Painting relation, but in order to facilitate multiple labels for any one painting, a weak entity ObjectLabel will be created, with a composite primary key of ObjectID and DateUpdated. GenreID will also be included in the Painting table, although it was required by only 23% of respondents in the survey.

At this point the refined schema representing Works of art and relationships include 11 relations (Table 5.3)

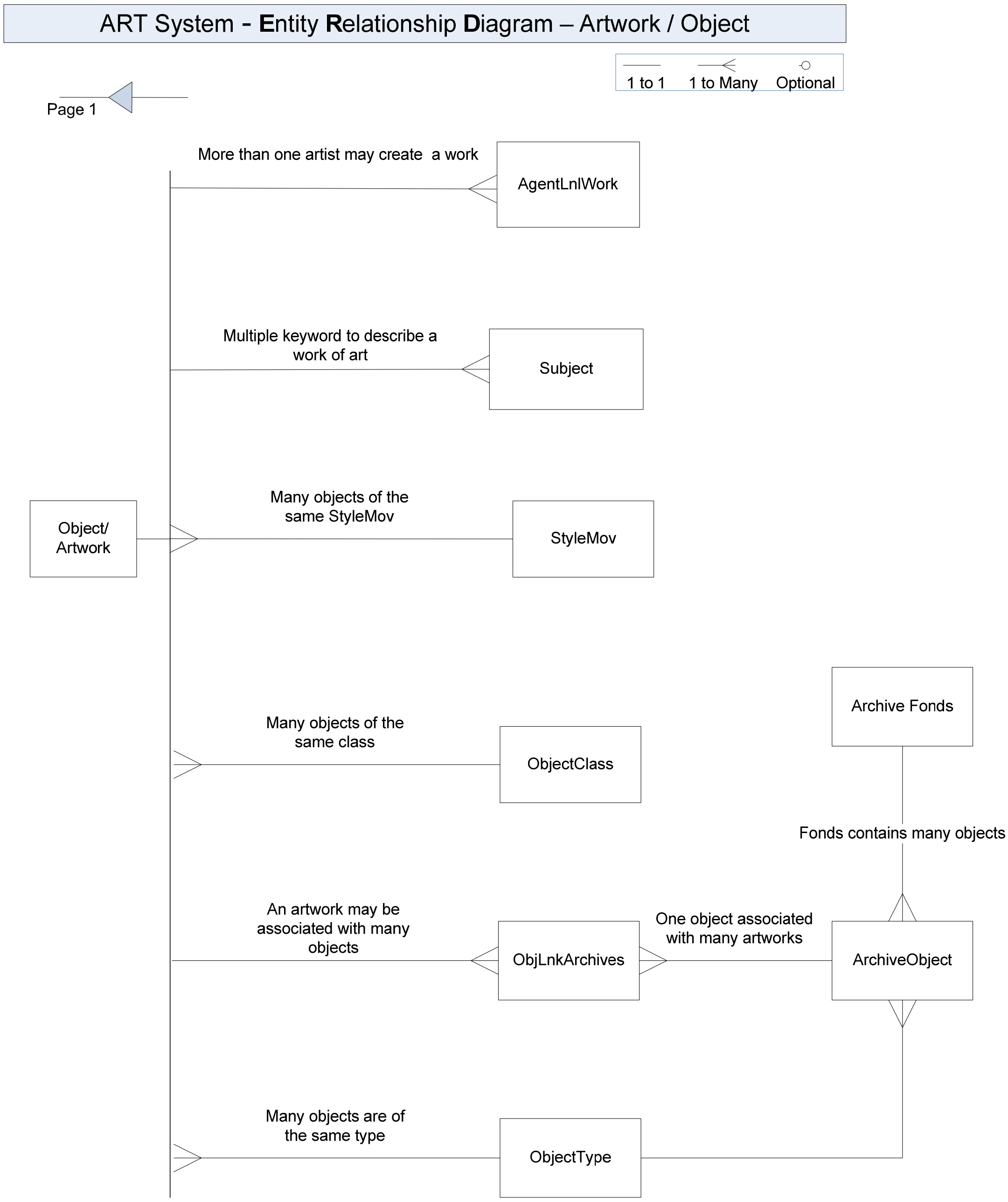


Figure 5.2 Entity Relationship Diagram (ERD) for Object/Artwork and associated relations

5.2.2.3. Relationships between objects – Artwork-Artwork / Artwork-Archives

Currently, establishing relations between objects such as a painting and another painting may be achieved using elements included in the metadata standards developed for the museum sector. One method of achieving this is by simply using a textual description of the related objects, in the main object record (Baca & Harpring, 2014 -Revised online 2016, p. 29). By way of example, the authors recommend inputting a brief description and credit line, that identifies the related work and distinguishes it from similar works "*Ruskin Hours; prayer book; unknown French; ca. 1300; J. Paul Getty Museum (Los Angeles, California, United States); MS. LUDWIG IX 3*). The authors also state that the kind of relationship between the work being described and another work should be input, for example, "*study for*, *preparatory for*, *part of*").

Alternatively, one or more specific metadata elements may be used solely to establish a relationship to another object and to describe the relationship between them. However, employing this method assumes that the cataloger knows the exact reference; the URI or file path, to the related object. Additionally, in both examples above, the referred object is unaware of the relation between it and the original object, lacking an automatically stored reciprocal relation stored in the referred object (although some standards, such as VRA Core 4.0, recommend specifying reciprocal relations where possible). Therefore, when viewing an object, for example a sketch in the archives, the curator is unaware of possible relations between that sketch and specific paintings. Inherent limitations of both of these methods however are the preclusion of explicit elements when relating to objects in another database, for example, multiple comments from multiple curators and curator element and date elements on which the curator contributions may be selected or sorted.

5.3. System Design and Implementation

The prototype web-based system we have designed comprises three layers, a presentation layer consisting of a user interface to interact with the database, the application/business logic layer to process user input and the back-end database services layer that interacts with the databases.

The overall system architecture is described (5.3.1) followed by a description of the graphical user interface (5.3.2). Formative evaluation received from art experts is the subject of the next section (5.3.3). System implementation and the technologies employed are discussed (5.3.4).

5.3.1. System Architecture

The system is designed as a web based application, based on a three-tier architecture model (Figure 5.3 Overview of system architecture

) comprising the following tiers or components:

A. The presentation layer is comprised of a number of web pages. Each page is presented in the web browser containing a graphical user interface that accepts requests from the user and formats the retrieved data for display on the page. The functionality is summarized to show that it allows the user to:

   a) Search and retrieve of works of art and their collaterals from the archives

   b) Establish a relationship between a work of art and materials in the archives

B. The middle tier is essentially an intermediary between the presentation layer and the back-end databases that:

   a) Processes query requests from the user and validates the data and return data retrieved from the server to the user

b) Sends the request (query parameters) to the back end database layer and receives the retrieved data from the back-end database layer, in case of a Select statement, whereas data is not retrieved for an Insert statement.

C. The back-end comprises the logic that builds the SQL DML (Structured Query Language Data Manipulation Language) queries and sends the SQL to the databases. This essentially is the inference engine where the work of art and all of its related entities are retrieved based on the SQL statements passed to the database

a) The first database stores the works of art and all relations except those with the archives.

b) The second database stores all of the archival materials.

c) The third database stores the established relationships between a work of art and collaterals from the archives.

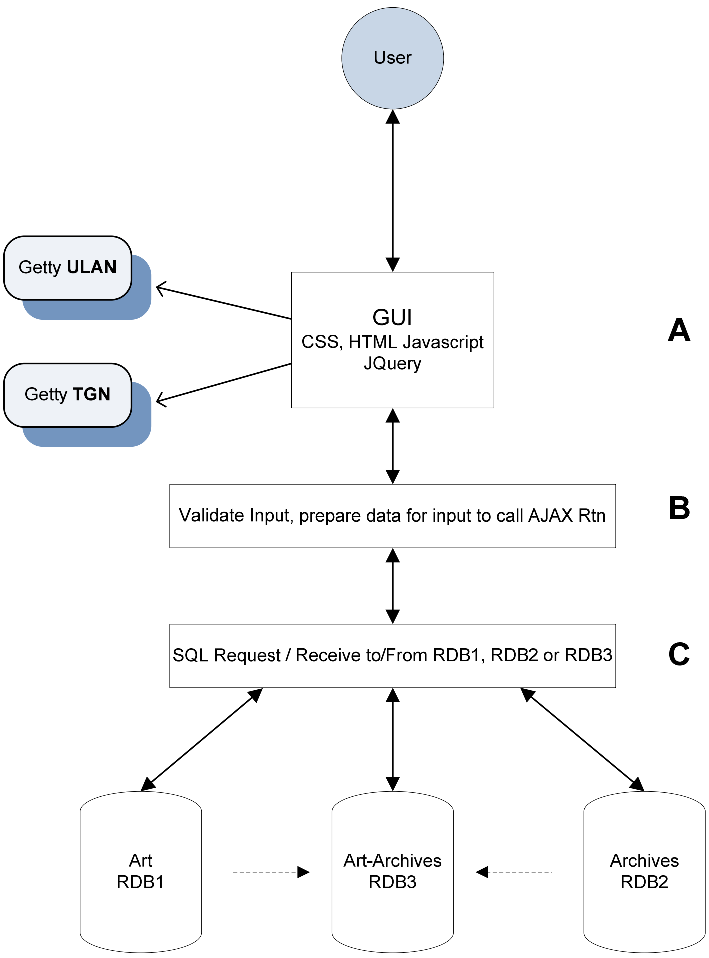


Figure 5.3 Overview of system architecture

The system components comprise A) the Presentation layer comprises a web interface from which the user may search, retrieve and visualize images of works of art and their collaterals from the archives and also create a relationship between a painting and materials from the archives. B) The intermediary layer that processes requests from the user before sending to the backend database and received the data sent from the databases. C) The back-end comprising three databases DB1 comprising all information about the aesthetic objects and their relationships, except for relations to archives, DB3 in the centre contains the relationships between the aesthetic objects and archival materials and DB2 stores the archives (positioned on the right to highlight the separation from DB1 and that the central database containing the relationships, sits between both databases.)

The components of the architecture and their inter relationships are illustrated in Figure 5.3 and these are briefly elaborated before discussing the user interface and system implementation details in the following sections.

A) Front-End

The front-end provides a number of options to retrieve data (and images). The user may do a simple search by choosing the artwork from a list of artworks, or from thumbnails of artworks. Alternatively, the user may elect to do an advanced search, by inputting or selecting search options based on the structured metadata stored in the database. On loading the advanced search form the system populates the dropdown lists with the relevant data from the database (for example, the object type to be retrieved, the technique and the supporting medium). When the user submits the form, the selected data are passed to the middle layer for processing and then to back-end where the query is built and submitted to the database. The retrieved data (comprising the work of art and all collaterals, including previously associated archival materials) is passed back through the same pathway, for visualization at the front-end. The user may also create a relationship between a work of art and archival materials and submit this for insertion into the link database, DB3

Supplementary information stored in the Getty vocabularies may be viewed by activating a hyperlink on the page, for example ULAN, to automatically display the ULAN record in a new window.

B) Middle tier

The middle tier processes the request from the front-end and validates the data, for example the latest date must be empty or greater than the earliest date (if that was specified). The middle tier passes the query parameters to the back end and on receipt of a response it forward the retrieved data to the front end.

C) Back-end

The back end constructs SQL queries based on parameters received from the middle tier and executes the queries on the database. Retrieved data is sent to the middle tier.

The back-end consists of three databases, DB1 comprising all information about the aesthetic objects and their relationships, except for archival materials. The database was populated using SQL scripts to insert data into the tables for the works of art and all related tables and links to the images, which are stored external to the database. DB3 in the centre contains the established relationships between the aesthetic objects and archival materials and DB2 stores the archives (positioned on the right to highlight the separation from DB1 and that the central database containing the relationships, sits between both databases.) The contents of DB2 (Archives database) were harvested from source text documents (Section 3.6), that describe the archives according to ISAD(G) (General International Standards for Archival Description). Data collection methods and database population methods are further described (0)

5.3.2. User Interface Design

The objective of the prototype system is to allow users to access the images and the information about the images that we have stored in our representation, underpinning an application. The system is designed as a web application that allows the user to visualize a work of art and its collaterals, including tombstone data and a description of the work of art. Unlike the typical visualization presented in museum systems and online, this visualization also includes archival materials generated and accumulated by the artist over the course of a lifetime. Among these documents and collectible items we may also find materials that relate to particular paintings. By way of example, one may find a preparatory sketch for a painting, or a photograph of a painting in an exhibition, or notes in the artist's diary relating to a painting or a letter from the artist's benefactor, or any number of things. These materials may also provide essential evidence about the provenance of the work of art. Each and every one of these unique and contemporaneous objects, and also assembled a whole collection can contribute to our knowledge of the artist and his paintings and the world in which he lived and weight of the archives in contributing to enhanced intellectual value is witnessed by the increasing interest in artist archives (Breakell, 2015; Pyle, 2001; M. Reed, 2017; Sjöholm, 2014).

Findings from the survey presented in Chapter 4 show that a large majority of the art specialists surveyed, consider archives in the top four most informative resources when seeking to build their knowledge of the collections in their care. Artist's archives can provide the missing background and contextual information that fills the knowledge gap and engenders deeper insight into a work of art.

Three options provide access to the images and collateral information. The first option is by selecting an artwork from a list of artworks, the second option allows the user to select an image from a gallery of image thumbnails. The final option allows the user to specify criteria for a restricted or more focused search. These options are described below.

The '**List'** option retrieves details for each image in the database and presents the information in a table (Figure 5.4). The system displays the total number of retrieved images at the bottom of the window and the ordering criteria at the top of the list. The user may scroll through the list to find the required image, and double click on a row selects the image and associated collateral for display in the 'show artwork' page. To assist the user in finding an image, the list may be reordered, by artwork title, earliest year of creation (the default sort order), the latest year of creation and by the artist identifier. Moreover, the theory underpinning this option of previewing the collection in a list that may be sorted based on these different parameters, is, that it may be of interest to the art specialist when curating an exhibition. By way of example, the exhibition may be based on a particular artist in the collection or it may focus on a specific time span, as discussed with a curator during one of our visits to the gallery.

List of Artworks: in ascending order of Artwork Title

TITLE ↓	EARLIEST YEAR	LATEST YEAR	ARTIST ID	ARTWORK ID	ARTIST NMAE	ARTIST BIRTH YEAR	ARTIST DEATH YEAR	TECHNIQUE	TECHNIQUE
(National Airs)Patriotic Airs	1923	1923	1	12	Jack Butler Yeats	1871	1957	Oil	Canvas
A Lake Regatta	1923	1923	1	7	Jack Butler Yeats	1871	1957	Oil	Panel
A Travelling Circus	1906	1906	1	28	Jack Butler Yeats	1871	1957	Watercolour	Paper
About to write a Letter	1935	1935	1	13	Jack Butler Yeats	1871	1957	Oil	Canvas
Above the Fair	1946	1946	1	5	Jack Butler Yeats	1871	1957	Oil	Canvas
Alone	1944	1944	1	40	Jack Butler Yeats	1871	1957	Oil	Canvas
An Occupation	1910	1910	1	30	Jack Butler Yeats	1871	1957	Oil	Canvas
Bachelor's Walk, In Memory	1915	1915	1	14	Jack Butler Yeats	1871	1957	Oil	Canvas
Before The Start	1915	1915	1	9	Jack Butler Yeats	1871	1957	Oil	Canvas
Bound for the Valley	1949	1949	1	23	Jack Butler Yeats	1871	1957	Oil	Canvas
Chair-O-Plains	1926	1926	1	36	Jack Butler Yeats	1871	1957	Oil	Canvas

Number of artworks listed: 47 (Double click on a row to view the artwork.)

Figure 5.4 Retrieve images details in list format

The user may reorder the list by Title, Earliest Year, and Artist and select an image by double clicking the row.

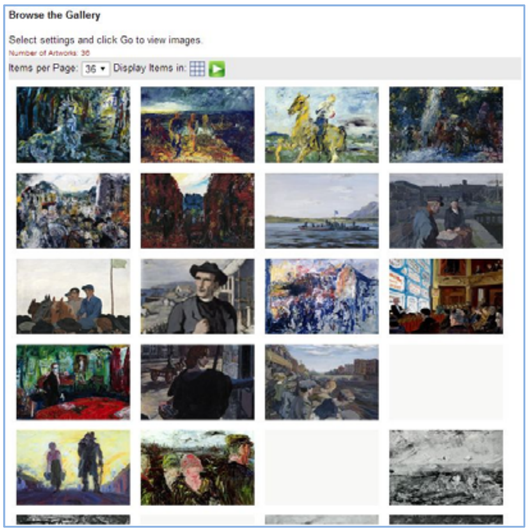


Figure 5.5 Retrieve thumbnail images

The user may select the number of images to retrieve from a dropdown list of preset number and select an image by double clicking on the image

When the user chooses the '**Browse Gallery**' option from the menu, the system retrieves thumbnail images representing the artworks in the collection (Figure 5.5). The only intervention allowed by the user is to select the number of images to retrieve, from a dropdown list of preset numbers. Presenting a view comprising the thumbnail images may be helpful to the curator planning an exhibition as the images are of sufficient size to discern the intrinsic qualities, theme, style and colours of the different kinds of artworks held in the collection. Double clicking on an image selects the image and associated collateral for display in the 'show artwork' page.

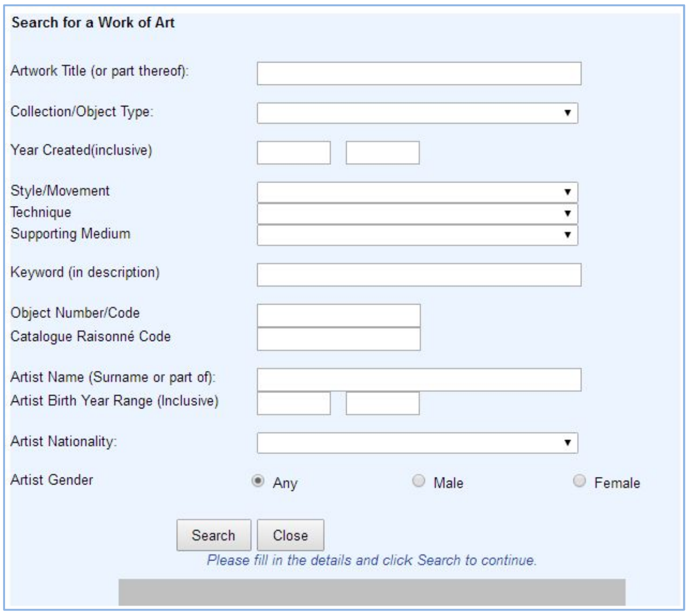


Figure 5.6 Advanced Search for a work of art

The '**Search Artwork**' option allows the user to perform a focused search of the artworks in the collection (Figure 5.6). The search parameters provided represent the classes of information and some of the properties or attributes of the data. The classes of data stored in the database comprise, the artwork, artwork/object type, Style/Movement, Technique, Supporting medium, Artist and Nationality. However, in this search for a work of art the main class is the artwork table and each of the other criteria are attributes of the artwork, including the artist. Having said that, if any of the artist attributes are specified such as Name, Male or Female, or earliest/latest year and nationality, then, this also limits the set of artworks based on the limited set of artists that may or may not satisfy the criteria.

The artworks retrieved must match all of the criteria specified, with two exceptions. If the user inputs either the Object Number/Code then all other selection criteria are ignored. Next, the system checks if the Catalogue Raisonné has been input and if it has then all other inputs are ignored. The reason for this is that the system assumes that the user knows these codes and if they are found then all other criteria are unnecessary.

The **show artwork** function presents (Figure 5.7) a large format image of the artwork, the tombstone data, a description and collateral from the archives if archival materials are associated with the image.

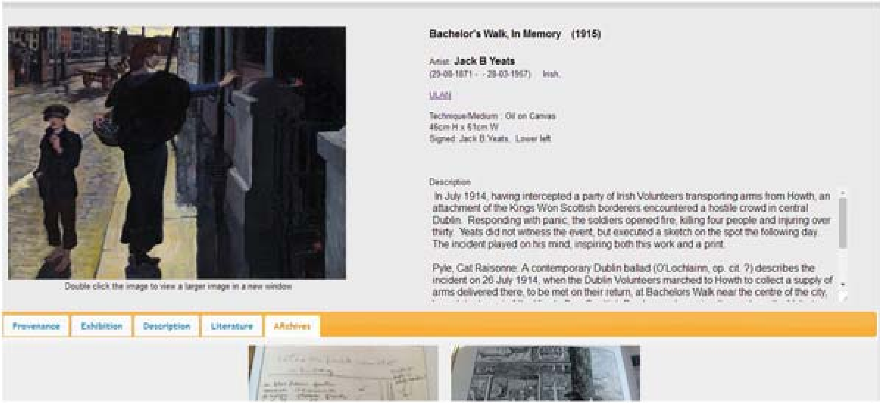


Figure 5.7 Show Artwork function

The zoom artwork function allows the user to change the change the size of the large format image of the artwork (Figure 5.8). The zoom archive function allows the user to increase or decrease the image for further inspection **(**Figure 5.9)

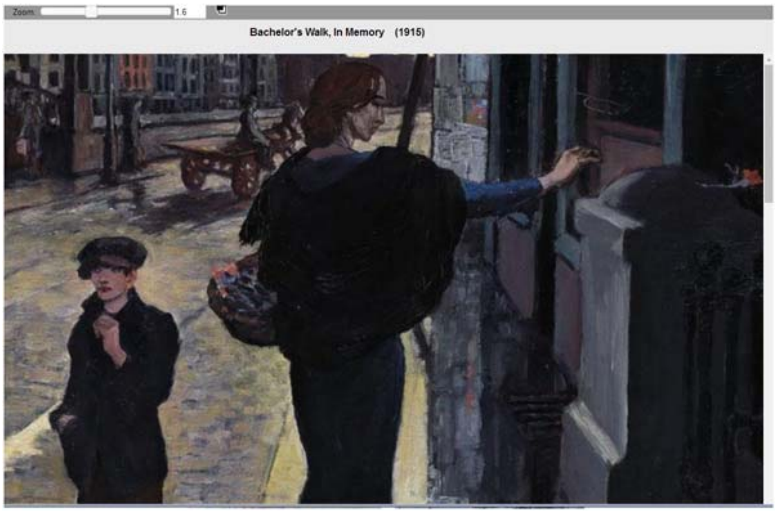


Figure 5.8 Zoom artwork (Up/Down)

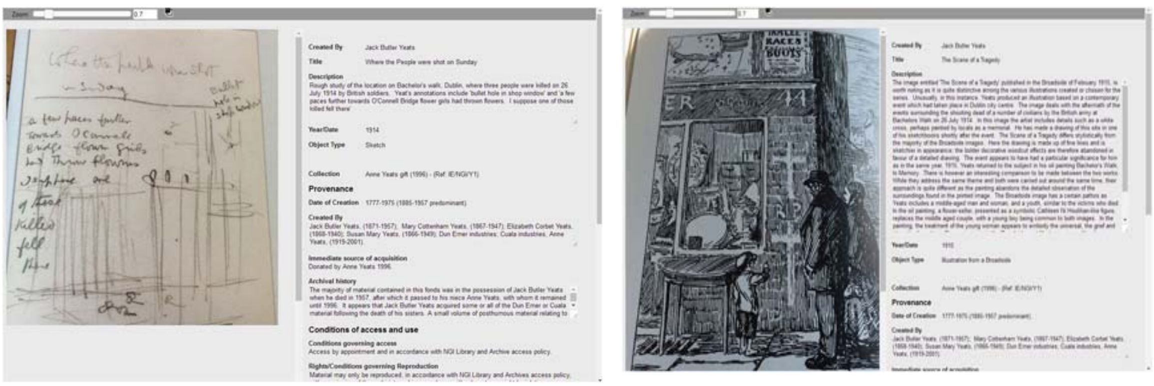


Figure 5.9: Zoom archive object up/down

Figure 5.10: Search the archives

Figure 5.11 List Archives

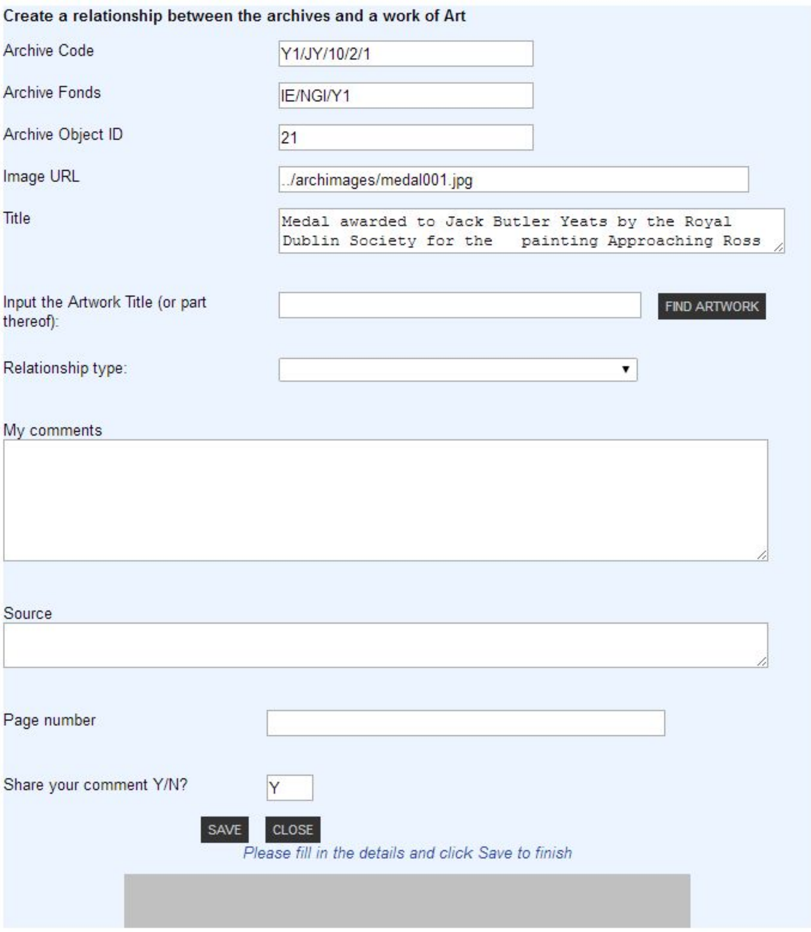


Figure 5.12 Relate an archive object to a work of art

5.3.3. Technologies

The design of a system that relates an aesthetic object to its collaterals description required a number of technologies to be used. Some would be very familiar to the reader and yet others would need a brief explanation (JSON). This section enumerates the technologies I have used to build the system (Table 5.5). The table shows the name of the technology, a brief description and the possible uses.

Table 5.5 Technologies used to develop the system

No.	Name	Description	Usage
1	Apache HTTP Server V 2.2	This is a free open-source Web server.	Installing the server on a laptop facilitates local development and testing of a web application.
2	PHP 5.2.11	PHP Hypertext Pre-processor is a free open-source server-side programming language.	Used to retrieve data from the database- by building and executing SQL SELECT queries. Used to add new (INSERT) data when harvesting archives from texts and when curator creating relationships between archival materials and works of art) to the database- by building and executing 'Insert' queries.
3	MySQL 5.1 SQL	Community version of MySQL (My Structured Query Language): an open-source relational database management system (RDBMS) available under GNU General Public License.	All data for the system are stored in the database. SQL Scripts were designed to create and populate tables in the database. SQL queries are also embedded in php programs to search, retrieve and add new data to the databases.
4	HTML	Hypertext Mark-up Language: is the standard mark-up language used to create web pages. HTML elements consist of tags enclosed in angle brackets. Source: ttp://en.wikipedia.org/wiki/HTML	Used to create and mark up web pages
5	JavaScript	JavaScript is a dynamic computer programming language. Source:http://en.wikipedia.org/wiki/JavaScript	In this application it is used in the Web browser to handle user input, run client-side scripts to interact with the user, control display the browser, initiate calls to communicate asynchronously and alter document content display.
6	AJAX & jQuery	AJAX (Asynchronous JavaScript and XML): A group of interrelated Web development techniques used on the client-side to create asynchronous Web applications. jQuery: a library of JavaScript functions available under free license from MIT.	It enables background retrieval of data from the database and images from the server, without having to refresh a web page. AJAX is Implemented in this application using jQuery Jquery library of functions were used to detect and process user interactions such as double clicking to select an image, clicking a button to close a window.
7	JSON	JavaScript Object Notation: An open standard format that uses human-readable text to transmit data objects consisting of attribute–value pairs. Source: http://en.wikipedia.org/wiki/JSON	Used primarily to transmit data between a server and web application, as an alternative to XML. Data retrieved from the DB is encoded in this format, using php functions to do this..
8	CSS	Cascading Style Sheets:	A language to describe the format/display of content on HTML pages.
9	Dreamweaver	Software to edit web pages.	Software to edit web pages.
10	Adobe Photoshop CS4	Digital image editor	Used to manipulate artwork images, to create thumbnail images of the art work .jpeg, to adjust images to same size for Start page on the website, adjust images for display on single page

I have demonstrated how a mixture of well understood technologies can be used in both in conjunction with some of the very basic and well used technologies and some recent ones to develop a system where the user will have semi-automated access not only to look at an image of an aesthetic object like in the medium gallery systems, but that they can look at it further by providing semi automated links between the object and its collaterals.

The work at Trinity so far had initially concentrated on an ontology description of an aesthetic object, and the representation of such objects using prototypical ontology language. That work related to the image description.

My work is an extension of that, in that, an arbitrary number of objects can be linked to the aesthetic object using conventional database technology. The question arises as to whether we can use ontology language in particular, and other knowledge representation schema to do that. This question can be explored in substantial detail by using a system such as I have presented, as it provides robust and stable systems for such enterprises.

The design of a system that relates an aesthetic object to its collaterals description required a number of technologies to be used. Some would be very familiar to the reader and yet others would need a brief explanation (JSON). This section enumerates the technologies I have used to build the system, described above (Table 5.5). The table shows the name of the technology, a brief description and the possible uses.

JSON (JavaScript Object Notation) is designed as a human readable data interchange language and provides an alternative to XML (Extensible Markup Language). It encodes data for transfer between for example, a server and a server or from a server to a client, using key value pairs, denoting the variable name and its value between curly brackets, separated by a colon. For example, "Artistname": "Jack B Yeats". This data may also be encoded into an array (using square brackets) and nested arrays. Research shows that JSON provides a more efficient serialization of data than XML (Nurseitov, Paulson, Reynolds, & Izurieta, 2009; Severance, 2012). For this system, data retrieved from the database is encoded in JSON format before returning the data back to the middle tier for processing.

5.3.4. Database build

The database was created using the phpMyAdmin free software tool and SQL scripts were coded to build the tables and views in the database schema. SQL scripts were also coded to automatically populate tables in the database. These scripts were coded using Data Manipulation Language (DML), a subset of SQL used to insert, update and delete data in the database. Scripting the database facilitates regeneration of the database on the same or other computer at any stage. Sample SQL scripts can be found in Appendix G.

5.4. Data Acquisition and Database Population

This section describes the data acquisition, data preparation and data population of the database with data describing the artworks, the artwork images and the archival record and images.

5.4.1. Data selection – Works Of Art

The sources for the works of art included:

- National Gallery of Ireland website
- Jack B. Yeats, A Catalogue Raisonné of the Oil Paintings, by Hilary Pyle
- One complete exhibition from an Exhibition Catalog – Masquerade & Spectacle:The Circus and the Travelling Fair in the work of Jack B. Yeats, by Róisín Kennedy

Opportunistic sampling and random sampling methods were used in the selection of artworks for the database.

Opportunistic sampling refers to the selection of 15 Yeats artworks from the *Masterpieces from the Collection* exhibition, at the National Gallery of Ireland. This selection is justified by the fact that art experts at the gallery have intimate knowledge of these works and can assist with data definition and population, if required. A second opportunistic and focused sample comprised a complete exhibition entitled, 'Masquerade & Spectacle: The Circus and the Travelling Fair in the work of Jack B. Yeats', containing 23 artworks (Kennedy, 2007) and all were selected for inclusion even though some of the artworks are watercolours.

A *random sample* of 10 artworks was selected from the Catalogue Raisonné of Yeats oil paintings (Pyle, 1992), using ranbetween() Excel

Selecting paintings by other artists, assisted with more comprehensive testing. These included, 'A Connemara Village', a painting by Irish artist Paul Henry (1876-1958) a contemporary of Jack B Yeats, provides a complementary perspective on Ireland, particularly the west of Ireland, the subject of many paintings for both of these artists. It also highlights the very different style of painting of both artists. A painting by Vermeer 'Woman Writing a Letter, with her Maid' was also selected to facilitate more comprehensive testing of search functionality.

Data elements for all of the artworks selected were typed into a spreadsheet. Descriptions of the artworks were either copied from the web into the spreadsheet, or typed in manually from descriptions found in exhibition documentation or in the Catalogue Raisonné.

Images of the artworks were copied from the web and scanned from the *Masquerade* exhibition catalogue. For cases where images could not be located, the librarian in Early Printed books scanned the images from the Catalogue Raisonné and emailed them to me.

The artists selected for inclusion included Jack B Yeats, Paul Henry, Johannes Vermeer and Rembrandt. Paintings were selected for all artists except Rembrandt. This will facilitate more comprehensive testing also.

The images were pre-processed using Adobe Photoshop software and thumb nail images were also generated.

5.4.2. Data selection – Archival Materials

The researcher made several visit to the Yeats archives at the National Gallery of Ireland, during which the photographs of the materials were taken. Photographs were taken using an iPhone and this accounts for the somewhat inferior quality and resolution required for the images of the artworks in the system. The images were uploaded to a laptop where they were organized in an archives folder and named according to the content, as described in the archival texts. Descriptions of these materials was already provided when the records describing the archives were harvested from the source text documents (0)

The images were pre-processed using Adobe Photoshop software and thumb nail images were also generated.

5.4.3. Database population

Scripts were coded to automatically populate tables in the database. Data from the spreadsheet of aesthetic objects was copied into SQL scripts, to facilitate automatic population of the database. Similarly, scripts were created to populate the database with the artists and the archival images. These scripts were coded using Data Manipulation Language (DML) DML refers to a subset of SQL and is used to insert and update data in the tables or to delete data from tables. Scripting database population facilitates regeneration of the data at will along with portability of the database. An example population script is included at Appendix G.

5.5. System Evaluation

System evaluation was carried out by the researcher running experiments (queries) on the database. The researcher organized meetings with experts at the National Gallery of Ireland.

The system was demonstrated on an individual basis with three experts in their offices at the National Gallery of Ireland. Each evaluation session lasted approximately 45 minutes and comprised 3 parts. For the first part the researcher gave a brief description of the system and the underlying theory. The next part comprised demonstrating the system to the expert and the third part comprised discussions and the researcher took notes of their comments and feedback, which were summarized to each one at the end. Feedback from the art experts, noted during the demonstration and in our discussion after the demonstration are summarised and categorized under: Concept, Functionality, User Interface

Concept: The three experts responded favourably to the concept of showing archival materials with a painting (Table 5.6). However a number of issues emerged in relation to copyright, data sharing and storage of the data and these are addressed in the table. Copyright issues are outside the scope of the study. The researcher is aware of copyright and while the system includes paintings by the Irish artist Jack B Yeats (1871-1957), who died less than seventy years ago, the system is only available on the researcher's laptop and is not available on the web. Also, recall that licences were purchased for images in the survey images.

Table 5.6 Evaluation of the system Concept

<table>
<tr><th>1</th><th>Concept</th><th>Questions and statements</th><th>A</th><th>B</th><th>C</th></tr>
<tr><td>1</td><td colspan="2">Show archives together with the work of art</td><td>3</td><td>2</td><td>2</td></tr>
<tr><td colspan="6">
Question: (A) Are the objects in the same databases or other databases?
Answer: No. Different databases. They are different types of objects and many museums keep them in different systems and databases so I tried to replicate that. Some museums may keep them in the same databases.
Question: (A) Can other images (of artworks) link to the artwork?
Answer: At the moment no, it is not programmed to do that. Technically possibly future work.
Question: (B) (i) Is it possible not to show comment I put in when I relate an artwork with an object in the archives? (because I might not want to share it until I'm finished my research);
(ii) Can only experts use the system or can anybody user it? (Her concern was that experts would only have confidence in the system if they were sure they were viewing comments from other expert (and not novices).
Answer: To use the systems at the moment you must log in to it. So anybody logged in can make a comment. Perhaps, if we let anybody view it but only allow users who are logged in to establish relationships and to input comments. This is technically possible.
Question (C): What about the copyright of the images?
Answer: That was outside the scope of the research. Users are expected to be copyright holders unless they say otherwise.
Question (C): Where are the comments stored? - Will the system try to update NGI record?
Answer: The comments and links are stored in a database for this application and it does not try to update any other system
</td></tr>
</table>

Functionality: The system provides a range of functionality to search and retrieve objects from the databases, to display the objects and to link archives and artworks. Evaluation of the functionality (Table 5.7) provided shows that while the functions fulfill the role, varying opinions among the experts emerged regarding missing data and placement of data and these are addressed in the table.

User Interface: Each of the windows provided in the system was evaluated (Table 5.8) with respect to the content, layout and the style. Standard text boxes, combo boxes or dropdown lists, radio buttons and textareas (larger the boxes that can be expanded by dragging the bottom left corner) were used in the systems and the associated labels for these elements were positioned to the left of the elements, except for the textarea, where the label was placed above the textarea. Evaluation of the user interface reflects the fact that this is a prototype system and the researcher and the evaluators acknowledge that a completed user system needs a professionally designed interface requires professional graphical skills.

Table 5.7 Evaluation of the system Functionality

2	Functionality	Questions and statements	A	B	C
1	**List Artworks**		3	2	3
	Question (B) Why does it show two years (referring to earliest year column and latest year column)? Answer: In case they are different - I noted when viewing exhibitions that sometimes the exact year is not known, and sometimes a ? Is shown and sometimes a From/To year is shown Re-ordering the list was viewed as helpful but see also UI evaluation.				
			A	**B**	**C**
2	**Artwork Thumbnails**	**Questions and statements**	2	2	1
	Question: (A) (B): Can we see more data besides the Title in the tooltip, like the artist name and the year of the artwork? Answer: Yes, that is possible, in future work **Question: (C): Can we see the Title, Year and artist name under the image? (that means you don't have to move the mouse over each one to find out the these details)** Answer: No. it is not programmed to do that, but it is possible and will take a bit of time, but can be done in future work.				
			A	**B**	**C**
3	**Search Artworks**	**Questions and statements**	3	2	3
	Question Expert (B): Is it possible to search by 'Acquisition type'? (That would be important (or example, to search for all Donated, Bequests, Purchases by particular funds) Answer: Yes, that is technically possible. Possible future work, and we would need to agree where exactly that should be shown. **Question Expert (B): Why is Catalogue Raisonné there, I don't think that is needed and we don't have that?** Answer: I have it captured for most of the paintings in the system, but it can be removed if it is not needed.				
			A	**B**	**C**
4	**Display Artwork**	**Questions and statements**	2	2	2
	Question Expert (B) (C): Is it possible to show the 'Acquisition type'? Answer: Yes, that is technically possible. Possible future work, and we would need to agree where exactly that should be shown. **Question Expert (A)(B)(C): Is it possible to the see Accession number?** Answer: Yes. That would be possible, In a future update				

Evaluation of the system Functionality (continued.)

<table>
<tr><td></td><td></td><td></td><td>A</td><td>B</td><td>C</td></tr>
<tr><td>5</td><td>Zoom Artwork/ Associated Archive object</td><td>Questions and statements</td><td>3</td><td>3</td><td>3</td></tr>
<tr><td colspan="6">Comment (A)(B), (C): Commented on (i) the poor quality of the images and (ii) to ask what was the black/white icon.

Response: To explain, the Images for the paintings were either downloaded from the web or in cases where not found on the web then from an exhibition catalogue using a phone, and/or scanned by the duty librarian in Early Printed Books at Trinity College. The images for the archives were taken on my phone during visits to the CSIA

The black/white icon allows the user to view a black and white version of the image (software generated).</td></tr>
<tr><td></td><td></td><td></td><td>A</td><td>B</td><td>C</td></tr>
<tr><td>6</td><td>Show Archives (List)</td><td>Questions and statements</td><td>3</td><td>2</td><td>3</td></tr>
<tr><td colspan="6">Question Expert (B): Can we see these in a Tree view also?

Answer: No. it is not programmed to do that. For future work I can search for a library tool, maybe there is one that enables that easily.

Question Expert (C): I wonder should we show the name also? - no I see it in the fonds row, so that is ok.

Answer/comment: Yes, the title of the fonds included the creator so that is why I didn't put in a specific column for the creator, and it leaves more space for the title/description.</td></tr>
<tr><td></td><td></td><td></td><td>A</td><td>B</td><td>C</td></tr>
<tr><td>7</td><td>Show Archives (Record)</td><td>Questions and statements</td><td>2</td><td>2</td><td>2</td></tr>
<tr><td colspan="6">Question Expert (A)(B)(C): Why show all the blank boxes?

Answer: I thought it would be best to show all (26) available fields rather than just the ones that contained data, because then it would be obvious which ones where not filled in. But that is an easy thing to change. (Note: The researcher changed the program to hide the title and the box if there is no associated data in the database)</td></tr>
<tr><td></td><td></td><td></td><td>A</td><td>B</td><td>C</td></tr>
<tr><td>8</td><td>Link Archive-Artwork</td><td>Questions and statements</td><td>2</td><td>2</td><td>2</td></tr>
<tr><td colspan="6">Question Expert (A)(C): How many words can be input, is it a short sentence?

Answer: For now I set it at 5000 characters, but this can be reviewed

Question/Comment: Expert (A)(C) It seemed OK but they need time to think about it.

Response: No problem, for now it is programmed to allow you to input but not to change or delete afterwards. That's something to think about.

Note: Expert B had already commented in the Concept section about perhaps holding off sharing data until the research was finished. Experts A and C needed to think more about the option.</td></tr>
</table>

Table 5.8 Evaluation of the system UI (User Interface)

<table>
<tr><th>3</th><th>User Interface</th><th>Questions and statements</th><th>A</th><th>B</th><th>C</th></tr>
<tr><td>1</td><td>List Artworks</td><td></td><td>2</td><td>1</td><td>2</td></tr>
<tr><td></td><td colspan="5">Question (A,B,C) The text at the top and bottom of the table is informative but it is hard to read because it is too small. Can the size be increased?

Answer: Yes, the font size is small and can enlarged in the CSS

Question (B) (C) (i) The grey heading of the table is a bit dark and it is hard to read the column titles (ii) can the column title be changed to normal case (not all capitals?)

Answer: Yes, that can be changed in the CSS</td></tr>
<tr><td></td><td></td><td></td><td>A</td><td>B</td><td>C</td></tr>
<tr><td>2</td><td>Artwork Thumbnails</td><td>Questions and statements</td><td>2</td><td>2</td><td>2</td></tr>
<tr><td></td><td colspan="5">Question (A,B,C) The text at the top of the window is too small. Can the size be increased?

Answer: Yes, the font size is small and can be easily enlarged in the CSS</td></tr>
<tr><td></td><td></td><td></td><td>A</td><td>B</td><td>C</td></tr>
<tr><td>3</td><td>Search Artworks</td><td>Questions and statements</td><td>2</td><td>2</td><td>3</td></tr>
<tr><td></td><td colspan="5">Question Expert (B): Do the buttons for gender need to be so spread out/ Can they be moved closer together?

Answer: Yes, that can be done

Question Expert (A): Is there a particular reason why the Search and Clear buttons are not in the centre?

Answer: No, I will take a look at that to see why they are not.</td></tr>
<tr><td></td><td></td><td></td><td>A</td><td>B</td><td>C</td></tr>
<tr><td>4</td><td>Display Artwork</td><td>Questions and statements</td><td>2</td><td>2</td><td>2</td></tr>
<tr><td></td><td colspan="5">Question Expert (B): I would prefer not to have those tabs at the bottom because you have to click each one to see the information. Would it be possible to show all the information under the description?

Answer: Yes, the tabs can be removed in a future update. Would it be better then to show the archive thumbnail images under the main image?

Expert (B): Yes, that would be best, and maybe they don't need to be that big

Question Expert (A)(C): I don't think the archive images need to be that big, half the size or less and they would still be visible and especially if there were more related materials, you would have to scroll down to see the thumbnails. Is that possible?

Answer: Yes. That is true, and that is possible</td></tr>
</table>

Evaluation of the system UI User Interface (continued.)

			A	B	C
5	**Zoom Artwork/ Associated Archive object**	**Questions and statements**	3	3	3

Comment (A)(B), (C): The layout is fine but the image quality is bad.

Response: Yes, as noted previously the Images were not taken by a professional photographer. The researcher either downloaded them from the web or took photographs using a phone or were scanned by scanned by the duty librarian in Early Printed Books at Trinity College. The images for the archives were taken on my phone during visits to the CSIA

			A	B	C
6	**Show Archives (List)**	**Questions and statements**	3	2	2

Question Expert (B): I know I mentioned it before but can we change the layout to a Tree view or maybe have both options?

Answer: For now, no, it is not programmed to do that. For future work I can search for a library tool, maybe there is one that enables that easily.

Question (B) (i) That looks like the same design as the List Artworks - the grey heading of the table is a bit dark and it is hard to read the column titles (ii) can the column titles be changed to normal case (not all capitals?)

Answer: Yes, that can be changed in the CSS

Question (C) (i) The grey heading of the table looks the same as the List Artworks but its easier to read here - maybe because of the wider column - (ii) I think it would be best to be consistent and make the same changes as to the List artworks table?

Answer: Yes, that can be changed in the CSS

			A	B	C
7	**Show Archives (Record)**	**Questions and statements**	2	2	2

Question Expert (A)(B)(C): You are going to get rid of the blank boxes?

Answer: Yes. I will
(Note: The researcher changed the program to hide the title and the box if there is no associated data in the database)

			A	B	C
8	**Link Archive-Artwork**	**Questions and statements**	3	3	2

Question/Comment: Expert (A) (i) Why are these buttons white on black and and the others are black on grey? (ii) are they different because you can enter data at this screen?

Response: No. They should really all be the same. I am using a B13different form so I think the CSS for that is different. I will make them all the same in a future update.

5.6. Summary

The representation and the implemented system described in this thesis is informed by art experts needs previously identified in the interviews and the survey (Chapter 4.)The bases of our representation that incorporates the knowledge of the art expert comprised a relational database schema that integrates long established metadata standards developed specially for the domain. These metadata standards assisted in organizing and describing the entities of interest and enabling relationships between these entities. Employing tools and techniques developed by database experts provided the methodology to design a robust database schema for this representation. These tools comprised conceptual modeling and logical modeling using the entity relationship model and the extended relationship model, followed by functional dependencies and normalization techniques to design a well structured schema that represented aspects of the real world. We have demonstrated that relational database technologies are well suited to represent the important entities in the domain and the many complex interrelationships between works of art and collateral materials.

When conducting their research of the collections in their care, art specialists consult many sources to build their knowledge of the artworks and they regard primary sources of artist archives as rich resources for essential background and contextual information to gain insight into the authentic nature of the artwork. Consequently, the challenge for this study was to design a representation that combines these primary sources with works of art to support these specialists when conducting their research of a work of art.

The system design comprises a web application, underpinned by the representation and it provides the art specialist access to the representation through search, retrieve and display functions that shows the work of art and related objects from the archives. The user may also establish relationships between the collateral objects and works of art and record their reflections about the relationship.

Database creation and population is achieved by SQL scripts that initially create each of the required tables and views and further scripts are coded to populate the tables with the artworks and their collateral information. Population of records to capture the archives required a custom written program that harvested archives from text documents (described according to the International Standards for Archival Description ISAD(G)). Subsequently, examination of the archives records and the archives images identified the matching records and SQL scripts were coded to record these relationships.

Formative evaluation from 2 experts during the interviews ensured that the user was aware of the source archive in which an object (for example, a sketch) resided, when an object was displayed with a related work of art. Evaluation of the completed system by 3 art experts revealed that although the system functionality required some minor changes, the user interface needed amending to achieve a professionally designed and consistent interface throughout the application.

6. Conclusion

This research was motivated by the question as to how to access materials from collaterals to an aesthetic object and perhaps more ambitiously to display the collateral materials together with the aesthetic object in question. This objective has been accomplished to a point that it led on to the other questions either in relation to representations and to accessibility. Recall the painting "The Liffey Swim" (Figure 6.1) by the artist Jack B Yeats, for which he won a silver Olympic medal and the key collaterals to the image include the medal, the register of awards and the aide memoire that suggested that Yeats had in a handwritten index of his paintings. This being 1924, having a typed document is a gift to the future digital researcher that could be scanned promptly, but the handwritten document would really involve handwriting-recognition software.

Figure 6.1: The "The Liffey Swim" (centre), a painting by Irish artist Jack B Yeats (1871-1957)

The Liffey Swim: winner of the Silver Olympic Medal (left) in the Arts & Culture section, at the 1924 Paris Olympics. Also shown, is a Register of Awards (top right) and a page from a Register of Paintings (bottom right) from the Yeats archives, at the National Gallery of Ireland

Both of these collateral documents have been accessed manually by researchers for many years and provide insight into how Yeats worked, lived and painted. Working closely with curators and archivists to talk about the implementation of the collection of the archive documents, revealed that that this was the mainstay of their work. Meetings with a curator at the Chester Beatty Library verified the import of archival materials. Both the National Gallery of Ireland (NGI) and the Chester Beatty Library (CBL) are National Cultural Institutions, one started as a private collection (CBL) and NGI was established in 1864. After Irish Independence NGI had an enhanced role for reference works by major Irish artists and this involved not only preserving the work of art, but also anything that was collateral to the work of art. Hence, there are warehouses of aesthetic objects and collateral materials, from reviews from critics, to artist responses, to bills and to brushes and sometimes even clothing belonging to the artist.

To me, this was a highly interconnected system of objects and their descriptions. Managed manually with the help of a dedicated curator and archivist this system is accessed by students of history, art history and culture.

Due to space limitations in the gallery, it is not possible to display these valuable enlightening and often humorous and tragic objects. However, institutions like NGI have inventory systems where descriptions of such objects are kept in database format. My interest was to explore and exploit the inventory which is not closely collateral to an object, it is not a strict or a precise description of the painting or the painter, as is the case in generally available digital collections such as, the 40,000 Meisterwerke [49]. However, modeling of such disparate objects is a commonplace problem in database design. One only has to go to the supermarket to see an instance of such a model where shelves are seldom empty and goods are coordinated centrally from warehouses hundreds of miles away.

Large state enterprises also use conventional databases produced by large companies such as Oracle and IBM. I have been professionally involved with systems analysis and database design for many years, so I have some awareness of this problem.

The key challenge here was to relate the inventory to high level standards, produced by museums and cultural heritage (2.3), where metadata and vocabularies enable consistent description and interoperability. These cannot be mapped automatically as it is a major knowledge engineering problem. The major standards and vocabularies motivated the design of the system for art and collateral objects. As the fundamentals of database design of aesthetic objects are seldom discussed I have explained these (3.2 - 3.5).

The collateral description of the aesthetic objects and the archives were available in texts. Reviewing the standards of encoding these texts for archival materials and the role that standard terminology plays in that, informed the design of an automated harvesting process that populated the database schema describing archival collections, from these source texts (3.6)

With a draft model of aesthetic objects and their collaterals in place, the next step involved collecting data from the experts, which is costly in terms of time and resources. Nevertheless, I conducted knowledge acquisition sessions, comprising face-to-face interviews with 4 experts in Ireland, who between them had about 45 years experience in the field and an online survey that yielded valid responses from 48 art experts, representing 806 years experience (4.3.6.1). This assisted in further refinement of the model and the metadata description of a work of art and its collaterals. These tasks, combined with meetings, visits to the gallery and archives and reviewing standards developed for the domain, allowed me to exam the content of primary sources and the art expert descriptions (4.2 and 4.3). This confirmed the initial hypothesis that materials are distributed in terms of their locations and are disparate in terms of adherence to

[49] The Yorck Project, Berlin, 2007 (last viewed 22 April 2018)

standards of description with some variation in description, confirmed by expert survey and interviews, Chapter (4.)

The knowledge of standards for aesthetic objects and experiential knowledge of archivists creating descriptions of collateral objects and the metadata standards for these collateral objects together with knowledge of how to design a database, led me to design a prototype system, comprising a data repository and a web application (Chapter 5.

6.1. Research questions revisited

The principal research question was:

How can a computational representation of a complex and knowledge rich domain support research in that domain leading to enhanced intellectual value?

This research is located in the cultural heritage domain and a selection of paintings by the Irish artist Jack B Yeats serves as the test case. In order to answer this question the following three sub questions were identified.

1. *What kinds of information are considered by art specialists when researching a work of art?*

Creating a good representation presents an interesting challenge given the complexity of aesthetic objects such as paintings. Designing a representation that incorporates the knowledge of art experts requires an understanding of the domain and the important aspects of the domain from the art expert's perspective. The literature review highlighted the complex nature of aesthetic objects and emphasised the need for experienced art specialists to preserve the value of these aesthetic objects. For that reason we contacted art experts at the National Gallery of Ireland (NGI), hence this study is carried out in cooperation with experts at the gallery.

The literature review also showed an increasing interest in artist's archives as potentially rich resources that may provide background and contextual information about the artist and the artwork.

In order to answer the question I carried out a case study that included interviews with 4 curators: three curators from NGI and one curator from the Chester Beatty Library, offering a complementary perspective on the domain. These curators represent a combined total of 45 years experience. I also carried out a survey of 48 art specialists from 9 different countries, with a combined total of 806 years experience. Findings from the study showed that the provenance of an artwork is a primary concern for these specialists (along with the current location and a description of the artwork) but that associated materials from artist's archives represent the next most important resource for their research. When focusing specifically on artist archives 94% of respondents confirmed the value of these primary sources when conducting their research, where 56% strongly agreed and 38% agreed that artist archives can deepen understanding of the artist and the artworks.

2. *How can database design principles support an enhanced representation of art that may underpin future intelligent applications?*

Relational databases are used worldwide for almost 50 years and may seem to be an obvious choice, or perhaps not, based on the technologies we have today. However, given the complexity of this domain and the challenge of combining these two very different collections, of aesthetic objects and collateral materials from the archives, that may contain thousands of objects, we decided to investigate this established model. Database technologies are secure and robust and scalable and can handle vast amounts of structured data for easy access. However, to leverage the strengths of relational database technology requires utilising the tools and techniques developed by database experts. These tools provide the required methodology to design a database that accurately represents the real world entities and their interrelationships and that supports search and retrieval. Furthermore, incorporating metadata standards developed specifically for the domain facilitates organisation and description of the appropriate entities and their relations to represent works of art and collateral materials.

3. *How can the system inform and support domain expert's research process?*

This perhaps is a complex question and an ambitious one.

To answer this question I designed a web application underpinned by the database schema of our representation. I collected images of artworks and hundreds of images of archival materials, from the Yeats archives at NGI and carried out a number of experiments and queries on the database following which evaluation of the system was carried out with 3 art specialists on an individual basis. The art experts are well experienced in using technology and using advanced search windows poses no problem for them. Evaluation by the art experts revealed that each of the experts agreed with the concept of combining archival materials with aesthetic objects. Findings from the evaluation show that although the system functionality required some minor changes, the user interface needed some amendments to achieve a professionally designed and consistent interface throughout the application. This was not unexpected given that this is a prototype system.

6.2. Towards a more comprehensive representation

Evidence of the strong connection between aesthetic objects and artists' archives motivated the design of the prototype for this research. The literature highlighted the increasing interest in artist archives and the potential of these unique materials to engender deeper insight into the artist and the works of art (Breakell, 2015; Pyle, 2001; M. Reed, 2017; Schwartz & Cook, 2002; Sjöholm, 2014). Recent research also shows how archives can be integrated with collections of artworks (Koot, 2015), and that including collateral materials from the archives opens up the possibility of uncovering new knowledge regarding relationships between the artist, art dealers and museums (Koot & Koppenol, 2016). Findings from the mixed-methods study (Chapter 4.) confirmed that art specialists, responsible for collections in their care, consider that these primary sources, can offer a variety of perspectives in relation to a work of art (Table 4.9). Consequently, the design of the prototype incorporated both of these heterogeneous and complementary collections (artworks and archives), to support the art researcher and to provide a more comprehensive representation than that typically available.

Determining specifically how the prototype provided a more comprehensive representation than typically available, involved further evaluation and reflection (beyond previous testing and evaluation by art experts). By comparing the prototype with online museum systems, in conjunction with the literature and findings from the mixed method study, under the following headings, a) knowledge capture, b) search and retrieval and c) archives, a clearer picture of the true nature of the representation, emerged.

Due to practical limitations but also recognizing the need for a representative sample of museum systems other than only Irish museums, the researcher randomly selected ten museums (Table 7.2) for this review. The selection of museums comprised three Irish museums (The National Gallery of Ireland (NGI); The Hugh Lane (HL, Ireland) and the Irish Museum of Modern Art (IMMA) together with a further seven world renowned institutions: The National Gallery of Art in Washington, DC (NGA); the Museum of Modern Art (MOMA, New York); The Metropolitan Museum (MET, New York); the National Gallery UK (NGU), the Louvre (Paris, France) and the Van Gogh Museum in the Netherlands (VGN). Including the Art History Institute in the Netherlands, (RKD) (previously identified in the literature review) ensured the availability of artist archives online, for this review. Finally, it was not possible to display and examine an abundance of artworks in each museum, nonetheless reducing artwork display to a maximum of three artworks in each museum afforded discovery of more types of knowledge represented, than by displaying only one artwork.

6.2.1. Knowledge Capture

Comparison between the prototype and the online museum systems, (excluding archives, which are analyzed separately) demonstrated a high degree of similarity in relation to the kinds of knowledge captured, shown in tabular format (Table 7.2.). Consistent with the literature (2.2) the majority of the systems (including the prototype) incorporated essential knowledge beyond that intrinsic to the artwork, to contextual details of the “who, what, where, why and when”. The systems presented tombstone data together with a description of the artwork, and where minimal information about the artist was displayed, the museums provided access to more details. Similarly, the prototype supplemented the limited information with biographical and bibliographical details, accessed via a hyperlink to ULAN (Union List of Artist Names, Getty Vocabulary). Half the museums omitted provenance of the artwork, despite the significance of provenance from both the aesthetic and legal perspectives (2.2.2), and top place ranking by (96%) survey respondents (Figure 4.8). Furthermore, less than half the systems incorporated exhibition history and bibliographic details, also previously identified by both the literature and the mixed methods study. The prototype, however, included provenance, exhibition history and bibliographic details related to an artwork, albeit for only a small number of the artworks.

Additionally, almost half the museums hold only a subset of the knowledge items listed (Table 7.2) for each artwork. By way of example, NGA included audio resources and infrared images for Vermeer's 'Woman Holding a Balance', but not for all artworks in the collection.

According to previous studies (pg. 41), image comparison is central to the art research process and recent evidence from the survey, concurs: 'images online' represented the most frequently used resources, by 83% of respondents on a daily basis, and ranked as the 'most highly informative resources' by 60% of respondents. Additionally, 60% of respondents considered the relationships with other artworks important to their research. However, when displaying an artwork, only four museums automatically incorporated related artworks, either presented as images or by textual hyperlinks (for example, a text hyperlink presented with the artwork entitled 'A Woman seated at a Virginal', NGU). By contrast, six museums and the prototype offered no indication of related artworks when displaying an artwork.

Finally, only three museums displayed keywords associated with an artwork: twenty two Iconclass notations (Figure 2.2) described a Vermeer painting at NGA; three keywords described 'Portrait of Lena Milius' at RKD; and VGN presented seven keywords next to the painting, 'The Potato Eaters'. In the latter case, however, the keywords did not necessarily describe the painting, but served more as an aid to continue searching the collection. This lack of keywords is consistent with the limited interest demonstrated by survey participants (Figure 4.8) and the researcher concluded that perhaps the respondents overlooked the benefit of these for search and retrieval. The remaining seven museums and the prototype omitted keywords, when displaying a specific artwork.

6.2.2. Search and retrieval

Comparison between the prototype and the online systems revealed a high degree of similarity in relation to the Search options (Table 7.3), but uncovered fundamental differences in relation to precision and recall. A default search in the online museums facilitated a wildcard search encompassing multiple properties of the collections (although it was not always clear which properties the systems compared). More advanced search options provided by some museums, including the prototype, allowed the user to narrow the focus of the search.

Examination of each of the museums search criteria and filtering options, depicted in tabular form (Table 7.3), clearly shows that the search options available in the prototype exceeded those provided in seven of the online museums. In addition to a wildcard search of the artwork title, description and the artist name, the prototype enabled searching by one or a combination of multiple criteria (Figure 5.6.), which closely matched the advanced search criteria in the remaining three museums, NGI; NGA and RKD. Filtering options in the online systems compensated for the limited search options in some museums and the advanced filtering categories, available at four museums (NGA, NGI, VGN and RKD) further improved precision and recall.

Combining and visualizing archives with aesthetic objects is central to this study, however, only two of the online museums automatically recalled archival materials when searching for an artwork. When searching VGN for a painting entitled "The Potato Eaters" the system retrieved a letter (a scanned image) among the twenty four different objects retrieved, whereas, RKD presented a named archival collection in the Archives panel in response to a search for a painting "Portrait of Lena Milius". By contrast, the prototype deferred retrieving related materials until the user selected to display the required artwork or archival object.

Finally, and similar to seven museums, the lack of keywords for both aesthetic objects and collateral materials, and lack of descriptions of individual archival objects (similar to RKD), such as sketches and correspondences or transcription of correspondences, precluded the discovery of potentially relevant materials. Similar to the Louvre however, the lack of filtering options subsequent to the initial search, diminished precision and recall and assistance to a user navigating through large volumes of retrieved objects, in contrast to the extended filtering options provided by four museums.

6.2.3. Artist Archives

Turning now to the final and fundamental part of this review, this section focuses on artists' archives. In particular, it examines how the museums facilitated the combination and visualization of archival materials together with aesthetic objects, from the point of view of the user or art researcher.

Few of the museums in this review incorporated primary sources of artist archives in their representation of an artwork. The paucity of archives online, however, was not wholly unexpected and is consistent with finding from studies of art specialists showing that the absence of archives online features among the variety of problems reported (Pg. 42). This examination was therefore limited to four museums, shown in Table 7.2, and it revealed that both RKD and VGN integrated archival collections with aesthetic objects, whereas the MET and the LOUVRE *refer* to or describe archival materials, such as a letter or a preparatory sketch in other collections.

With the exception of RKD, it was not always clear whether or not archival materials were available. The difficulty for this researcher was in finding either the hyperlinks or references to the archival material, which were sometimes obscured among the surrounding texts describing the artwork and at other times were located in another section of the webpage altogether. By way of example, when viewing the painting "Irises" at the MET, the 'Related Objects' section at the bottom of the page included five paintings, with no mention of related archival objects. However, further inspection revealed that the 'References' section of the web page contained references to letters in the van Gogh archives. Alternatively, a hyperlink in the body of the text describing “The Potato Eaters” provided immediate access to the letter in the Van Gogh Letter Repository.

Both the RKD and VGN allowed the user to browse through the archival collection(s). VGN provided functionality to view scanned images of letters; to search (by multiple different options), display and sort the materials and to show both Dutch and English transcriptions of the letters. On the other hand, RKD presented the archives collection and allowed the user to navigate through the finding aids, that is, to drill down the collection, organized in a tree structure, to reveal the related materials. Additional functionality allowed the user to search the materials using free text keywords. Since individual objects within these archives were not always described (similar to the prototype) it appeared that a matching result set included all objects recorded at the same ‘level of description’, (for example, fonds; series; file, Figure 3.20) whose description matched the wildcard search.

Significantly, none of the online systems (included in this review) allowed the researcher to create a link between the artworks and archives or to input comments about related materials.

The prototype: Similar to RKD and VGN, the prototype also enables the user to view, to search, to sort and to navigate through the archival collections. Evidence from the literature and the survey findings shows that artist archives contain potentially valuable materials for art specialist research and it is probable that the art specialist will discover materials of interest, when browsing the archives. However, should the art specialist, uncover a sketch or a letter for instance, that sheds light on some aspect of an artwork, what options are available to the art specialists? In the case of RKD and VGN, the art specialist must make a mental note or write it down. That is, here is no mechanism for the art specialist to record his/her opinions and knowledge about an object and its relations to another object.

The notable distinction between the prototype and systems that incorporated aesthetic objects and archives (based on the four museums in this review) is that the prototype offers further support to the art specialist community. It allows the art specialist to contribute knowledge about existing relations between artworks and archives and to further extend that knowledge by creating new links between the materials. This support is facilitated by functionality and features that allow the art specialist:

- To explicitly create a link between an object in the archives and an artwork
- To name, or assign a term to the relationship between the objects
- To record reflections and opinions about the relationship between the linked materials
- To share this knowledge (the relations and reflections) among the art specialist community
- To automatically recall previously related objects when viewing a work of art

- To automatically (provide a list option) enabling the user to view a list of artworks related to the currently displayed archival materials

6.2.4. Conclusion

This review shows that the prototype matched half the museums under the heading of ‘knowledge capture’ while surpassing at least two other museums. Under the heading of ‘search and retrieval’ the prototype search options exceeded those provided by seven of the museums, although it is acknowledged that filtering options would further improve precision and recall. Arguably, deferring the retrieval of all objects related to artworks until the user selects to view a specific artwork may or may not suit an individual art researcher. The literature and the survey confirmed the value of archival materials for art research and it is the integration and visualization of these combined materials together with a) the knowledge captured and b) the search and retrieval functions, that presents a more typical representation. However, taking that typical representation in conjunction with extending the knowledge captured through establishing relations and sharing this knowledge, is what supports the claim that the prototype presents a *more comprehensive representation than typically available*.

Furthermore, the features associated with establishing relationships between materials, listed above, contribute to an ‘*enhanced representation’* which is underpinned by structures for efficient storage of each relation (once only) with links to the knowledge (comprising the reflections and the term assigned to the relationship) contributed from many different art specialists.

6.3. Findings

Turning now to more general aspects of this thesis, it is worth noting that this research provided an important opportunity to reassess and advance understanding and knowledge of art specialists needs in relation to preserving the collections in their care.

Generalization of findings: The number of respondents to the survey conducted in this study appears low, comprising only 48 valid responses, and this may preclude generalizing the findings to the population. However, all of the experts involved throughout this research, at meetings, when visiting the archives, together with the interviewees and the survey respondents, were highly qualified art specialists. The study respondents represented a combined total of approximately 850 years experience, from at least 9 different countries, working in 20 different museums, throughout Europe, USA and Canada.

Moreover, the findings from this study also corroborate the literature review that highlighted the increasing interest in artists’ archives and recent innovations in this area and contributes to the knowledge and substantive theory in the domain

Knowledge of the provenance of a work of art represents the primary concern for the majority of art specialist (96%), followed by publications (81%) and the location of an artwork and the artwork description. Evidence of the stronger connection between artist archives and works of art emerged, where 69% of respondents selected archival materials, slightly more than the 60% expressing interest in the relevance of other artworks to an artwork (60%). Moreover, when the question focused specifically on archives, 94% of survey respondents believe that artist archives can deepen understanding of the artist and the paintings and confirms how different perspectives emerge through the lens of these unique, contemporary materials generated during the artist’s lifetime.

Further analysis required: Given that 69% of respondents reported ‘Cataloging’ as a core function of their role, some surprising findings from the survey are worth mentioning. Only a limited number of respondents possess knowledge of the standards in the domain (13% claimed to know many of the standards while 64% reported that they had some knowledge). Just over half (54%) reported that they knew what standards were used in their institution. Allied to this is the unexpected limited interest in keywords. In addition to these findings, it should also be noted that the review in the previous section showed that only three of the museums used keywords to describe artworks.

Framework and prototype adaptability: The prototype in this study showed how different repositories can be integrated and visualized in a representation, (in the arts domain). However, the *framework* for this research, comprising the knowledge acquisition phase, the technical design phase and the application design and evaluation phases, can equally be adopted and adapted for other similar type research in different domains. Furthermore, implementing the three-tier architecture (for the prototype) facilitates change while minimizing amendments to either the front end interface or the business and perhaps more importantly the back-end storage, especially in relation to re-use for a different domain and/or as the underlying data model evolves or changes.

6.4. Limitations and future work

6.4.1. Limitations

The researcher acknowledges a number of limitations of this study.

Art Culture: The reader should bear in mind that the literature review, including the standards designed for the domain; all of the experts consulted during this study; the interviewees and the survey respondents, represented predominantly Western art culture and only one survey respondent, responsible for the preservation of Islamic collections, noted the lack of archival materials for these collections. Perhaps incorporating other art cultures and their associated kinds of aesthetic objects would have an impact on the kinds of resources art specialists consult when conducting their research and hence the findings and the prototype designed for this study.

Linking artworks and archives: The prototype enabled the art specialist/user to explicitly establish relationships between an archival object and an artwork and to input their reflections about the materials. In order to relate the materials the user must locate an object of interest (either by browsing or searching using exact or wildcard search of the archival record). A drawback of this approach can be the (likely) case where thousands of objects exist and the difficulty for the researcher in finding an object or recognizing possible connections between the materials. The ability to find archival objects was also hampered due to the lack of descriptions and properties of archival materials, since these were harvested automatically from the source texts available from the museum without further enrichment of the metadata. The potential for automatically linking archival object with artworks (or suggesting links) was not investigated and is left to future work.

6.4.2. Future Work

Evaluation of the prototype by the art specialists uncovered cosmetic changes in relation to the user interface besides other minor changes and these are excluded from this discussion. Two important areas uncovered during evaluation and that need to be changed before the system could be implemented in the community, concern knowledge sharing and navigating and visualizing archives.

Knowledge sharing/confidentiality: The art specialists valued the opportunity to make connections between artworks and archival objects, together with capturing their reflections about these relations, but they were concerned about sharing these insights immediately with the community. For example, if they were conducting research for an exhibition or future publication, then this knowledge must remain private until such time as the exhibition or the research is complete.

Navigating and Visualizing archives: Ordinarily, archives collections are presented in a hierarchical tree structure (as witnessed at RKD online) and they are also presented in lists or tables (as witnessed in Van Gogh Museum). The prototype also presented the archives in table format, but a tree structure was the preferred format by one of the evaluators. The significance of visualizing the archives cannot be underestimated, however, based on evidence from discussions during the case study interviews (Pg.141) with art experts and from this researchers own experience harvesting and visualizing archives, in conjunction with comments from at least one participant in relation to the 'overwhelming' nature of archives (Table 4.9) and previous studies.

Automatic Linking: The potential for automatically linking archival object with artworks (or suggesting links), offers significant benefits to the art specialist research process, by easing or virtually eliminating the burden of searching through thousands of objects. Furthermore, automatically linking the materials could potentially uncover new relationships, (immediately) undetected by the naked eye.

One approach to consider for automating the links may be to employ the relational database toolset comprising, tables normalized to 4NF for efficient storage of multi-valued values (Table 3.3) combined with Views, SQL Joins and LIKE statements, and necessarily multiple indexes on object properties to support fast retrieval, for testing purposes.

However, a more robust approach would involve a Topic Map, an ISO/IEC 13250 standard, previously described (2.1.6), extending linking capability to automate links between artworks and archives; between artworks and artworks; and between and among archival objects (beyond the means of the wildcard search currently provided in the prototype. Moreover, superimposing a Topic Map above these base systems and expanding this to reference other external documentation and systems in the domain would support discovery of additional related knowledge.

Nevertheless, successfully automating links between materials, regardless of the technologies selected, depends upon consistent description of each object. An object's properties would need to describe the object, for example, type; title; description; the author/creator of a letter, sketch or photograph; the recipient of a letter; the content of a letter, sketch or photograph; the titles of individual artworks contained in a letter or bill of sale and perhaps even the transcription of these documents. Similarly, the consistent application of keywords (for example, by using AAT, (Art and Architecture Thesaurus) to the artworks would enhance the ability to automatically link artworks to artworks.

Navigation and display of Archives and related objects: Viewing the archives at RKD, the Van Gogh Museum and the Delacroix Museum and NGI, exposed not only the lack of consistency in the properties presented, but also a lack of uniformity in display and navigation of the materials, making it difficult for a user to move from one collection to another. Future research regarding display and navigation of archival finding aids would be worthwhile, especially as more archives come online. For example, expanding navigation methods beyond the common approaches of drilling-down a hierarchical tree or paging up/down/next/previous to process an archives collection, while also enabling enlarged images of the materials (presented in a similar manner to multi-slice/multi-volume medical images) may simplify navigation and searches of archival materials, while preserving the context and access to related materials.

6.5. Contributions

This research convincingly demonstrates that knowledge management for art specialists needs to focus on linking works of art with collateral archival materials.

The design of a clear and sound representation of works of art and their collateral archival materials in the framework of an extended entity relationship model, informs future developments in the domain.

The proposed representation has been implemented using a straightforward relational database approach, and has resulted in a prototype system that was used to verify curator requirements.

This study also contributes to the body of knowledge in relation to the requirements of curators and art specialists when conducting research in art, as reported in the mixed-methods study.

6.6. Summary

The overarching aim of this research was to design a representation (solution) that supported art specialists' research in the knowledge rich domain of cultural heritage with a focus on museum curators, responsible for collections of aesthetic objects in their care. This thesis presented a framework that incorporated methods of knowledge acquisition, technical design and application design and evaluation, in order to achieve that aim.

The knowledge acquisition phase comprised a literature review that revealed an increasing interest in artist archives and highlighted the potential of these unique and contemporaneous materials for research in art. Furthermore, the research revealed a variety of implementations in the domain and a recent innovation showed how archives can be integrated with artworks with the possibility of uncovering new knowledge. Findings from the mixed-methods study comprising interviews with 4 curators from two national cultural heritage institutions in Ireland and 48 art specialists from 9 different countries worldwide (with a combined total of 806 years experience) corroborated the significance of artist archives for their research, where 94% of respondents confirmed that artist archives can deepen their understanding of the artist and the artworks.

The immense body of knowledge that continues to accumulate in the cultural heritage since the 1970's, relating to metadata, vocabularies and cataloguing guidelines, informed the method for harvesting archives from texts and populating the database. The knowledge gained from the mixed-methods study together with knowledge of the standards in the domain and ongoing research, in conjunction with methods of conceptual modeling and logical design, contributed to the design of a solution that demonstrated how value was added through these methods.

Evaluation of the prototype by art specialists verified the value of integrating artworks and archival materials and the art specialists found that the additional functionality to relate specific archival objects to an artwork together with recording their reflections about these relationships, helpful to their research. Notwithstanding the limitations identified above, this thesis provides a sound foundation for future research in this area.

7. Appendices

Appendix A

7.1.1. Interview Protocol

This appendix contains a scaled down version of the interview guide used for the semi-structured interviews with the art experts the case study, described in

	About You
1	What is your **role** at the museum?
2	**Experience:** How long are you working in this role? Previous experience?
3	**Qualifications**: What is your background / area of study/qualifications
4	In the course of your work **do you carry out research** of art works or assist others in their art research? For example, *perhaps for exhibition, research for conference, publication, education, other)*
	Curator: Functions and information about works of art
1	I have some idea about what a curator does, for example, curate exhibitions. **What functions are critical to curation of the collections**
2	**What *kinds* of information about work of art important to the curatorial process?** *(Prompts: For example- Acquisition (Funding, Provenance, Accession), Restoration – reason, when, by whom), Catalogue – ID, index for search, description, creator, date created, title), Exhibition – Theme, Title, time span, other? Artist – see ULAN + events)*
3	Aside from collection acquisition documentation w**hat other kinds of records and media are used for curation of the collections?** For example? 1. Database system 2. Paper based files 3. Books 4. Web based – website? 5. Web based – upload to Union Catalogues e.g. Europeana, Artstor? 6. Journal publications 7. Conference papers 8. Other
4	Besides the website: **what in-house computer systems are used for collections curation?** **Have you identified any limitations or shortfalls in these systems?** missing data? Can you establish relationships between artworks, between artworks and archives, other?
5	**Besides the in-house computer systems, what other computer systems do you use for collections curation?** *Prompts: For example, auction house sites, collection restoration, cataloguing standards, related thesauri*

	e.g. AAT in TMS), other, Art Library, Academic literature online?
6	**Standards for collection management**: Do you know what standards (metadata or vocabularies) for describing artworks such as paintings are used at your institution and/or are these standards automatically inbuilt to the software systems for collections curation?
	Resources for collection curation.
1	**What information resources do you use when curating Artworks such as paintings?** For example: 1. Art books 2. Consult with colleagues 3. Web based systems 4. Conferences – attend, participate 5. Journal articles 6. Visit exhibition at the gallery – attend, participate 7. Visit exhibition online – view, participate 8. Visit the Archives – physical or online 9. Visit Library – physical or online 10. In-house computer system 11. OTHER? Which do you use most frequently use / your preference?
2	**How frequently do you use web based systems when researching an artwork?** **NN** Times per: **Month, fortnight, week, day,**
3	**Are there particular websites that you use when researching an artwork?** Is there a particular reason for using these websites?
4	Concentrating now on **online** resources - **What types of online resources do you use when carrying out your research?** Museum websites? Academic Journals? Other? How would you rate your satisfaction with the online resources that you use, with respect to finding the information you are looking for? Excellent, Very Good, Good, Fair, Poor, Very poor What do you like or find particularly *helpful* when using online resources? What do you like least or find *problematic or least helpful* when using online resources?
1	**ARCHIVES** **Have you ever consulted the archives when researching an artwork?** *If Yes - Why so / If No - why not? If Yes – Did you visit the physical archives? And/or Online archives?*
2	In what way OR do you think that archival materials may / may not enhance intellectual value and representation of an artwork?
3	Following on from the previous question - **Are there particular kinds of materials from the archives that may support research into an artwork/painting? How so?**
4	**How to present archives with a painting?** Given the principle of 'respect des fonds' - What is the best way to show e.g. a sketch or a letter ? Does the viewer need to see all upper level descriptions? Does the viewer need to know from what collection a sketch is from, for example?

Appendix B

7.1.2. Coded Extracts from Interview Transcripts

ID	Interview text	Category	Code	Notes
A171	P: No. I'm not saying that they aren't, it's just that in general we, we don't know. I mean I know of course what you are talking about and I'm sort of making a joke by saying no, **because I think in general that most curators don't even know how this works**. And if they know they might be interested in it but they don't know how to work with it or	Object	metadata	Believes that most curators don't understand metadata standards or how to use them
A177	P: Yes, True, **but it doesn't mean that we always fill it in the same**, that we would always fill in the same thing in the same way. For example, this is something say with eh Provenance, is the way that eh there are no real as to how to write down Provenance. **There is no dictionary of Provenance writing**, so, or maybe there is one but nobody uses it. But em, how, so people write it in different ways and therefore it can also be interpreted in different ways.	Object	metadata	inconsistency in recording data. no 'dictionary of Provenance writing'. Different data can lead to different interpretations.
A179	P: No, but what happens is that generally standards, influential publications generally become sort of standards.	Object	metadata	Believes that standards may emerge from influential publications
A144	P: Well these is there is of course the archives.	Resources	archives	Other kinds of records would be ARCHIVES
B153	P: Yes, but the thing about the archives is that they can have a h**uge amount of information in them** that is very accessible because its text and people can read it. But it's, it's also not as, visually, I suppose **it's not as easy to access**, if you know what I mean, if that makes sense. So, putting archive letters on line, on a system, it can be quiet difficult for people to read, they **have to be a very high quality scan and very high resolution** and you need to be able to zoom in on things. That's why generally researchers will always like to consult the original, because it's just its easier to sit down and read a piece of text like that.	archives	archmat	requires good quality images and functionality
B155	P: Yes, no it is yes, but something like a handwritten letter you need a very high resolution eh version of it. You need to be able to read it. It's what makes putting archives on display or on websites challenging, that, it's not like putting a nice picture and people will or might want to see the details but they don't need to see it to get a sense of what's there. Whereas archives, you need to be able to see the detail to get anything from it. Otherwise you're just looking at some sort of eh, an abstract image.	archives	archmat	Challenge is the huge amount of materials in the archives and necessity of good quality images
C250		Archives	archmat	diary entry

Appendix B Coded Extracts from Interview Transcripts

ID	Interview text	Category	Code	Notes
C287	P: Yes. But I think you can do that in a very discrete way. Again, you don't want to overwhelm someone but you need to have it so that if someone is more curious about it that the information is easily accessible. So whether that's just a discrete link that will bring you to another website that might have a pdf of the full archive. Say for the Yeats archive the finding aids are available [pdf] online, but they are not in your face. So it's just about finding a way to show that's it's there if people want it, but they are not facing this onslaught of information.	Archives	archmat	not overwhelming user with all materials/ the complete archive. Rather, present one object and provide link to archive collection
A221	P: But this also means that for example we have Yeats archive here so if you make an exhibition of Yeats you'll spend a lot more time in the archives than I would do. If you would write a book on a former director of the National Gallery you could spend a few months here in the archives. You know, if you want to know, if you want to write about the history of the gallery then you would do that. If you want to make an exhibition of Vermeer you know you probably won't spend that much time...In the archives here.	Archives	archopin	depends on what you are working on. Good exhibition – gallery publication
A225	P: Yes it does. But it depends on individual cases. You know, how much. Em, because, Yes. They're just pieces of the puzzle. You know, that's it. Eh but again, if you're interested in the, let's say, the style of the painting, it means the history of the painting, the provenance and if you have a very thick file on the provenance of the painting, isn't very interesting if you are interested in the style of the painting.... that the problem.	Archives	archopin	pieces in the puzzle. BUT depends on the task in hand whether archives are relevant or not.
B147	P: Always, yes, it's essential really to use archives. You can't, as I said you can't, I mean there are historians who have, T J Clarke wrote a book about a Poussin painting that he stared at every day for, I can't remember if it was everyday for weeks or months, and wrote about the experience of what he was seeing, but it was more as an experimental piece of writing about art than art history and but you can't do or write art history without archives.	archives	archopin	Essential to use archives
B149	P: Absolutely, yes, well the same way as you would about any history. You need the contemporary perspective on it or the words of the author or the artist themselves.	archives	archopin	Contemporary perspective is needed for good Art Criticism or Art Appreciation
C248	P: Yes, if we can. If we can go in person, fantastic and if it looking at it online, yes.	Archives	archopin	Positive opinion

Appendix B Coded Extracts from Interview Transcripts

ID	Interview text	Category	Code	Notes
C281	P: Yes. Just to find out the history of the painting. So whether that's through Provenance or its visual history, say if it's based on sketches, if there's a diary entry that talks about it. Just to find out the history of it and what's going on behind the image, yes. I'm in the archives every day [here] but I would visit the National Library at least once a month. And in relation to particular kinds of archive materials, well it's whatever is there, I mean everything has some kind of use. So, whether its sketchbook, diaries, ephemera, books that the artist owned, photographs, usually whatever archival material is there you can find a use for it. It could be something like an installation shot from an exhibition in the 30's but it will show you how that painting was hung, what it was hung beside, maybe if there's someone important looking at it, everything, **every piece of an archive will tell you something about the artwork**. It all adds up.	ARchives	archopin	valuable source of context and background details of the artwork. Curator visits archives every day (at the gallery) and the N. Library every month.
C281		ARchives	archmat	It's whatever is there: - these materials can prompt curator questions or ideas about an artwork
D172	P: Yes. Primary resources are very important.	archives	archopin	Important
D192	P: Yes, context is very important.	ARchives	archopin	context
C080	P: yes, exactly I mean that's a big part of it as well. I mean it's often quieter in the summer when you know maybe people are away on holidays, students are away and that type of thing. But then we would have very busy periods you know usually just before Christmas and around kind of March April can be very busy when people are doing essays and that kind of thing. So it is, we never just have one thing on the go, there's always something else. So that was the first main project that I was with and then I was doing.	Task	assres	Prep. materials from the archives for students / researchers, -- busy periods
D030	P: I do, absolutely... but we don't have the acquisitions budget to be acquiring costly materials.	Resources	Auction	Use auction access and track similar materials
D032	P: Em yes... so if we're talking specifically about the Western manuscripts sold sometime... a lot of them would have been acquired by museums.. they will not come back onto the open market. Some were purchased by private individuals or dealers and in that case a lot of them would have been sort of eh dis-bound and sometimes, occasionally what happens is that a folio, or a miniature or a text page will come back on for sale. Eh obviously if you're a dealer you make a bit more money if selling that way.. But that's not new, it's how the Chester Beatty papyrus were sold and acquired by Beatty.... em, so buying an individual leaf would be actually a little bit more in our acquisitions budget than buying. Of course we would love to have more but at least we have some examples...	Resources	Auction	Use auction access and track similar materials

Appendix B

7.1.3. Sample from the Codebook used when analysing interview transcripts

	Code Category	Code	Description	Notes
	Resources			**Find out what resources are used, when used and the value of these to their role. You may use the specific name instead of the CODE if the participant names a specific resource**
7		oj	online journals	
8		omus	museum websites	
9		ospec	online art specialist resources	Typical values: Auction house -Artnet, Getty, Open Access images, RKD
10		concol	consult with colleagues/other experts	
11		manres	offline materials	Typical values: books, collection catalogues, exhibition catalogues, visit exhibitions, archives
12		system	museums own systems	Typical values: web, TMS, Embark
13		visexhib	visit external exhibitions	
	Objects			What different kinds of data or information domain experts consider important to build knowledge of the collection. Artwork data or Archive Data?
14		objdata	Tombstone data, keywords for search/discovery, Provenance, Object description, Label, Exhibition	Expect to uncover different kinds of data relevant: cataloguing, finding, describing, understanding - May use the specific attribute in coding e.g. Prov, label, desc, title
15		Coll	Collections	What kinds of collections is person responsible for?
16		metadata	kind of information needed	What metadata is important to describe artwork?
	Archives			Uncover the contribution of archives to intellectual value of an artwork
17		archopin	opinions about archives	Opinions of how relevant are archives to the understanding an artwork
18		archmat	Different kinds of archival materials	Typical value for kind of materials may be sketches / sketchbooks, letters, books, photographs
19		contrib	Contribution	Can archives enhance understanding and knowledge of an artwork? Enhanced intellectual value through...?

Appendix C

7.1.4. Survey Screenshots

This appendix contains screenshots of the 7 pages of the web based survey discussed in Section 4.3. Some of the screenshots are cropped at the end (Next button is removed) in order to fit on the page.

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Survey of Art Specialists

Page: 1 of 7

Role of the Curator

Q. 1
We understand that curators carry out a range of functions towards preservation of collections.

Can you let us know what are the core functions of your role together with any other tasks that you may carry out from time to time?
Please click in the Core or Other column or leave blank for tasks that you do not undertake.

	Core	**Other**
Acquisition	O	O
Cataloguing	O	O
Exhibition Design	O	O
Conservation of the collection	O	O
Contribute to Gallery publications	O	O
Research – building knowledge of the collection	O	O
Research – publication in academic journals	O	O
Attend Conferences	O	O
Participate at Conferences	O	O
Provide assistance to art researchers	O	O
Provide Education/Workshops to the public	O	O
Other (Please describe below)	O	O

Other: Please describe other functions you carry out.

You may type up to 2000 characters in this box. You can increase or reduce the box size by dragging the handle at the bottom right hand corner of this box. (Keep the left mouse button

Next

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Survey of Art Specialists

Page: 2 of 7

Information Resources

Q. 2 **How frequently do you use the following resources?**

Please click in the frequency cells to make your selections. Rarely indicates up to a few times a year

Whereas blank/empty indicates that you do not use a particular resource.

	Daily	Weekly	Monthly	Rarely
1. Images – online	○	○	○	○
2. Texts about images – online	○	○	○	○
3. Art Specialist Books – online	○	○	○	○
4. Art Specialist Books – hardcopy	○	○	○	○
5. Consult with colleagues	○	○	○	○
6. Journal articles – online	○	○	○	○
7. Journal articles – hardcopy	○	○	○	○
8. The collection management system of your institution	○	○	○	○
9. Hardcopy files at your institution	○	○	○	○
10. Websites of other museums	○	○	○	○
11. Visit exhibitions at other museums	○	○	○	○
12. Other (Please describe below)	○	○	○	○

Other: Please describe any other resources that you use.

You may type up to 2000 characters in this box. You can increase or reduce the box size by dragging the handle at

Q. 3 **Which of these (same) resources do you find especially informative?**

Please indicate the value of each resource to your role, ranging from 1 (the lowest) to 5 (the highest)

	1	2	3	4	5
1. Images – online	○	○	○	○	○
2. Texts about images – online	○	○	○	○	○
3. Art Specialist Books – online	○	○	○	○	○
4. Art Specialist Books – hardcopy	○	○	○	○	○
5. Consult with colleagues	○	○	○	○	○
6. Journal articles – online	○	○	○	○	○
7. Journal articles – hardcopy	○	○	○	○	○
8. The collection management system of your institution	○	○	○	○	○
9. Hardcopy files at your institution	○	○	○	○	○
10. Websites of other museums	○	○	○	○	○
11. Visit exhibitions at other museums	○	○	○	○	○
12. Other (Please describe below)	○	○	○	○	○

Other: Please describe any other resources that you find especially informative.

You may type up to 2000 characters in this box. You can increase or reduce the box size by dragging the handle at the

Appendix C Survey Screenshots continued

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Survey of Art Specialists

Page: 3 of 7

Information Resources

Q. 4 **What do you consider to be the key characteristics of online resources to support your role?**

For the key characteristics that you select, please indicate how useful you find each one – ranging from 1 (the lowest) to 5 (the highest).

	1	2	3	4	5
Image quality	O	O	O	O	O
Description of artwork	O	O	O	O	O
Functionality e.g. Zoom, Navigation	O	O	O	O	O
Hyperlinks to other resources	O	O	O	O	O
Aesthetics	O	O	O	O	O
Availability	O	O	O	O	O
Other (Please describe below)	O	O	O	O	O

Other: Please describe any other characteristics that you feel are essential or beneficial.

You may type up to 2000 characters in this box. You can increase or reduce the box size by dragging the

Q. 5 A range of art specific resources exist online, some of which are listed below.
Can you let us know what are your preferred online resources?

Please select only the resources that you use and indicate how strongly you feel that they inform your research - ranging from 1 (the lowest) up to 5 (the highest)

	1	2	3	4	5
1. Getty Research Institute	O	O	O	O	O
2. Art Index	O	O	O	O	O
3. Oxford Art online	O	O	O	O	O
4. Artstor	O	O	O	O	O
5. RKD – Netherlands Art History Institute	O	O	O	O	O
6. Google Arts & Culture website	O	O	O	O	O
7. Auction Houses	O	O	O	O	O
8. Art section in Newspaper / Magazines	O	O	O	O	O
9. Other (Please describe below)	O	O	O	O	O

Other: Please describe any other art specific resources that are beneficial to your role.

You may type up to 2000 characters in this box. You can increase or reduce the box size by dragging the

Appendix C Survey Screenshots continued

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Survey of Art Specialists

Page: 4 of 7

Standards in Cultural Heritage: Metadata and Vocabularies

We see Tombstone data (such as artwork Title, Year, Medium, Dimensions, Artist Name and Artist time span) when viewing an artwork such as a painting, on many museum websites.

Yet, standards developed within the cultural heritage domain include a variety of other data for administering and documenting objects in collections.

Some of these standards include: CDWA, SPECTRUM, ISAD(G), DC, VRA, EAD, CIDOC CRM, EDM, along with vocabularies such as Getty ULAN, AAT and TGN and a classification system such as ICONCLASS

The following questions refer to standards used in this domain.

Q. 6
Are you familiar with any or many of the standards listed above?

Please select an option that best indicates the number of these that you recognize and may have used

○ All ○ Many ○ Some ○ None ○ Other

Other: Please describe any other standards that you use. You may also comment on these and any other standards in the cultural heritage domain, if you wish.

You may type up to 2000 characters in this box. You can increase or reduce the box size by dragging the handle at the bottom right hand corner of this box. (Keep the left mouse button pressed down to drag.)

Q. 7
Do you know if your institution has adopted any of these standards for managing the collections, through the collections management software and manual recording systems?

○ Yes ○ No ○ Not sure

Previous Next

Appendix C Survey Screenshots continued

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Survey of Art Specialists

Page: 5 of 7

Metadata for objects in the collection

They Come, They Come (1936)
Painting: Oil on canvas
61cm High x 91.5 cm Wide

Artist: Jack B Yeats, Irish (1871 – 1957)
Signed: Jack B Yeats (lower right)

© Estate of Jack B Yeats, DACS London / IVARO dublin 2016

Click to view another painting

Q. 8
Aside from tombstone data to describe an object.....
What other information do you consider relevant to your research and knowledge of an artwork?

Please select any of the items below by clicking on the text

- [ ] Description
- [ ] Provenance
- [ ] Current location
- [ ] Artwork style / movement
- [ ] Genre
- [ ] Critical Reviews
- [ ] Publications about the Artist / Artwork
- [ ] Relationship to other artworks
- [ ] Keywords - theme
- [ ] Keywords - content
- [ ] Keywords - colour
- [ ] No. of copies
- [ ] Verso
- [ ] Associated Archival materials
- [ ] Relationships to other artists
- [ ] Other (Please describe below)

Other: Please describe any other data that you believe contributes to your scholarly research and knowledge of the collection.

You may type up to 2000 characters in this box. You can increase or reduce the box size by dragging the handle at the bottom right hand corner of this box. (Keep the left mouse button pressed down to drag.)

Previous Next

Appendix C Survey Screenshots continued

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Survey of Art Specialists

Page: 6 of 7

Considering Archives

Archives preserve a vast quantity and wide range of historical records. Among the archives we may find collections related to artists and/or to the artist's family.

Artist's archives may contain a variety of materials such as loose drawings, illustrations, sketchbooks, publications by the artist, correspondences, miscellaneous ephemera, the artists library, evidence of collaborations with other artists, legacy artistic materials and so forth.

We would like to know your views on the role of archival materials in research of the collection.

** The digital images below are available courtesy of the Getty Open Content Programme, the Open Access policy at the Rijksmuseum and the National Gallery of Art (Washington) and OpenGLAM **

Q. 9 **It is said that working with the artist's archives can deepen our understanding of the artist and his paintings.** Please indicate your level of agreement with the above statement by selecting one of the following options.

○ Strongly Disagree ○ Disagree ○ Neutral ○ Agree ○ Strongly Agree

Q 10 **In what way do you think that archives can contribute to enhanced intellectual value?**

Altering perspectives? New insights? Please let us know your views by typing in the text area below.

You may type up to 2000 characters in this box. You can increase or reduce the box size by dragging the handle at

Q. 11 **Have you ever visited the archives when researching an object in the collection?**

Please indicate how you accessed the archives, or not, by selecting one of the following options.

○ Yes – I have visited the physical archives and online archives
○ Yes – I have visited the physical archives
○ Yes – I have viewed the archives online
○ No – I have not visited the physical archives nor viewed them online
○ No – The archives I need are not online and I cannot travel

Appendix C Survey Screenshots continued

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Survey of Art Specialists

Page: 7 of 7

Demographics – all personal details will be anonymised

Q. 12

Age Range (Optional) ○ 18-25 ○ 26-35 ○ 36-45 ○ 46-55 ○ 56+

Q. 13 **Gender (Optional)** ○ Male ○ Female

Q. 14

Nationality (Optional)

Q. 15

Please select an option that best describes your role.

Q. 16

Recognising that roles change and that ylou may have performed different roles

How many years experience have you in your curation/art specialist profession?

Q. 17

Education - Major subject

Q. 18

What is your highest level of education in the selected subject?

Q. 19 **Email address (optional): If you will give further feedback please input your email address.**

Q. 20 **What is the name of the gallery, museum or institution that you are associated with (optional)?**

Submit

Appendix D

7.1.5. Meetings

This appendix contains minutes of some of the meetings with domain experts at the National Gallery of Ireland.

Meeting 1	10 October 2013 at the National Gallery of Ireland (NGI)
Attendees:	Dr. Adriaan Waiboer PhD, Curator of Northern European collection Prof. Khurshid Ahmad, TCD, Marian Reeves, PhD student, TCD
Purpose:	Introduction: Establish contact with NGI with a view to working in cooperation with art specialists, during our study. Gain insight into art specialist tasks and approaches to art.

We attended an introductory meeting with Adriaan Waiboer PhD, Curator of Northern European collection. (He has been working at NGI since 2004)

Initial discussions focused on Adriaan's recently published book entitled: *Gabriel Metsu Life and Work A Catalogue Raisonné,* 2012, as we had brought a copy of it along to the meeting. He recounted how he had spent many years researching Metsu; in fact the research started while reading a PhD in the US, in 2002? Based on the book: We enquired about the criteria used to group the paintings as we noted groupings called 'Search Aids' comprising following headings: Old Testament; New Testament; Saints; Literature; Various male professions and activities and so forth. We also asked about the sequencing of the artworks. The Search Aids reflect content of the images and although he didn't explain the sequencing at the meeting he does address this in the introduction to the book stating that the study '*presents Metsu's oeuvre in consecutive sections that examine groups of paintings that closely relate to each other in date of execution, style and subject matte*r'

We informed him that the focus of our research was Jack B Yeats. He pointed out that Metsu had created a limited number of works in comparison to Jack B Yeats who painted around 1200+ oil paintings alone. (His book documents 492 paintings classed as authentic (133), potentially authentic (21), known only through description (293), and rejected paintings (45)

He advised that perhaps we should limit our study, for example, to focus on a particular time span, theme or some other method, to make it more manageable and feasible to complete the study within the required time span.

We enquired about what tasks are carried out by curators and he informed us that he was very busy at the moment: in fact he is currently engaged in planning and organizing an exhibition of European art for 2014 at the gallery. When we asked how a curator might approach this task he informed us that exhibitions may be presented to celebrate a particular event or organized around a particular theme or group of artist or some other basis. He states that it can take a couple of years from the start of planning an exhibition through to actualization. He has identified artworks that he wishes to include in the exhibition (based on some theme – not explained to us). Acquiring the artworks may involve liaison with other galleries, for example, to get an artwork on loan from a gallery for the duration of the exhibition. He said that while this could be done by phone/other distance communication method that he prefers to meet with the curators; in fact he was due to fly to Germany the next day

He recommended that Dr Brendan Rooney & Janet McClean would be the best people to talk to about Jack B Yeats as they were responsible for Irish art and the Yeats collection.

The meeting concluded.

Meeting 2	28 January 2014 at the National Gallery of Ireland (NGI)
Attendees:	Catherine Sheridan, Assistant Librarian at NGI Leah Benson (LB), Assistant Librarian at NGI, Donal McGuire (DMG), at NGI Prof. Khurshid Ahmad, TCD, Marian Reeves, PhD student, TCD
Purpose:	Introduction: Establish contact Librarian & archivists with a view to working in cooperation with art specialists, during our study. Gain insight into archivists tasks and Yeats archives in particular

At a brief meeting with Catherine Sheridan (her background and education is in Library systems, she is also responsible for managing the Library systems at NGI.) she tells us that there are a number (how many?) of stand-alone computer systems for managing collections at NGI. This situation evolved because as new systems were implemented data was not migrated from legacy systems.

CS has to go but she introduced us to Leah Benson and shortly thereafter Donal McGuire joins us.

We had the rare opportunity of viewing recently acquired materials for the archives: this being the personal library and archives of Sir Denis Mahon. This vast collection, amassed over his lifetime, had recently been shipped to NGI and is housed in v large, mobile shelving. Acquiring this prestigious collection is recognized as an achievement and LB tells us how she (and another person from NGI) visited Sir Denis before his death, and that NGI was gifted the collection, over and above more prominent galleries.

Some of the tasks undertake by an archivist: LB explained that each and every item will have to be inspected and recorded (hand written on sheets of paper) to build an inventory of the materials. Specialist cleaning and storage requirements will be noted. The materials will be documented in detail after which they will be cleaned and conserved in specialist materials where necessary, before being organized into logical groups and moved to their final storage location. NGI do not currently have the personnel to complete these tasks and will have to take on new personnel. (On our way to the Yeats archive DMG shows us a piece of Harry Clarke stained glass – a lamp?)

Yeats Archive

This large and eclectic collection is housed in a room dedicated to the Yeats archive. Most of the materials are organized in covered boxes on trays and stored on shelves, in purpose built storage units, behind closed and locked doors. There is a label on each box or tray, containing a code. LB informed us that all archive materials are assigned a code and we ask about the Word documents on the website (there is one document for each of the 11 collections.) The codes on the boxes correspond to those in the word document.

We continue to view some of the materials after Prof. Ahmad leaves. I enquired about the Broad Sheets (described by Pyle and Rosenthal). LB procured a box containing (15?) Broadsheets. Each one is conserved in sealed/laminated clear sheet and the content is clearly visible for inspection. We also viewed the miniature stage that Yeats built (Hilary Pyle recounts how Yeats built a stage and put on plays for the neighbouring children, when living in the UK). We viewed other materials including a smock and paint brushes belonging to the artist.

A number of artworks (mainly watercolours) hang on the wall opposite the storage units and I recognized "Crystal Palace" 1903, as this was contained in the *Masquerade and Spectacle* exhibition at NGI (18/07/2007 to 11/11/2007): this exhibition has been selected for inclusion in the database created for this study.

LB informed that the archives include items relating to other members of the Yeats family and we viewed embroidery works by one the sisters (Lily or Lolly). These large pieces are framed behind glass and displayed next to the watercolours.

We took some photographs (see below.)

We agreed to meet again shortly and the meeting concluded.

Yeats Archive: Materials are stored in covered boxes. A label is placed on each box and a code denotes the Collection and position within the collection.

Yeats Archive: Sample of a Broad Sheet

Jack B Yeats commenced his artistic career as a graphic artist and illustrator.

While living in the UK he collaborated with artist Pamela Colman Smith to produce A Broad Sheet. A new Broad Sheet was created each month from January 1902 to December 1903. Each Broad Sheet was hand painted. Pamela Colman Smith worked with Yeats in 1902 only. Thereafter his wife Mary Cottenham Yeats worked on them.

The Broad Sheet was published by Elkin Mathews

Pyle, H. (1989). Jack B. Yeats, A Biography (2nd ed.). New York: Barnes and Noble Books.

Follow up (May 2014) On Inspection of the website in May I see that NGI are advertising job opportunities for personnel to work in the Archives and Library departments. Long term contracts (2 years?) are offered.

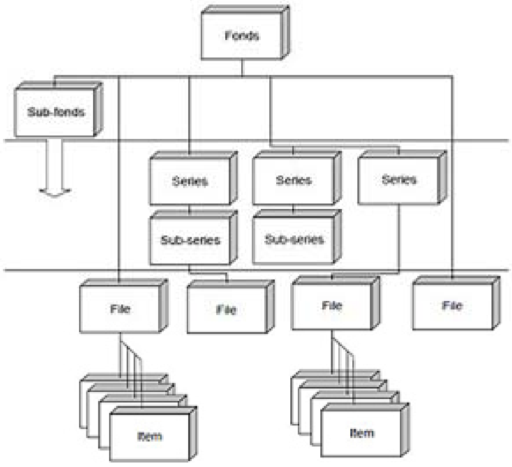
Follow up (October 2014) NGI have a new web page about the Sir Denis Mahon archives.

http://www.nationalgallery.ie/en/Research/GiftofKnowledge/Mahon/Mahon_Biography.aspx

At the end of the page we learn that “Between January and May 2010, an NGI team, including library, archive and conservation staff, organised and packed the entire contents. Close to 1000 bankers boxes were transported back from London to Merrion Square.”

Appendix D Meetings continued

Meeting 3	6th February 2014 at the National Gallery of Ireland (NGI)
Attendees:	Leah Benson (LB) & Marian Reeves (MR)
Purpose:	Learn what archivists do, how they handle materials and so forth



Archives personnel uphold the ***Respect des Fonds*** principle. That is, keep the archival records in the same order that they were created or in which you received them. Why? More authentic and maintains context. So, even if it might be easier to organize like objects with like – you cannot.

Rather than write down what we talked about in terms of the structure of archival materials– refer to the above picture.

That is the structure to be maintained in the archival entity.

LB says that archival materials received from Anne Yeats were perfectly organized and in good condition. Yeats books were first editions, Broadsheets were in good condition and needed to be conserved – specialist plastic sheeting.

Each unit of archives is kept as that unit – exactly in the order received. There are 11 different fonds in the Yeats collection, itemizing all materials in the archives: Yeats books (authored by him), sketch books, artist working materials – smock, brush and so forth.

Other issues such as privacy - 30 years restriction? preservation, storage. Archivists / she would probably be more strict than the 30 years - depending on the nature of the materials in the archives and the affect that opening them up to the public might have on possibly living relatives – whereas the affect on people's lives a few generations may be minimal.

We also talked about codes allocated to archives, codes allocated to art works, sculptures etc. IE/NGI/Y1, IE/NGI/Y2, IE/NGI/Y13, IE/NGI/26 – denoting country code, holder of the collection and Y for Yeats. And a number – not sure what the number represents as it does not run in sequential order.

Also, in the gallery's Yearly publications it seemed that a different set of numbers are used for paintings, sculptures, prints. LB explained that up to a few years ago that was the case but that now these separate groupings are not used; instead the next available number is allocated to the next acquisition irrespective of the kind of art work being acquired.

Aside: it used to be the case with the employee numbering system also... A set of employee numbers denoted Curators, librarians and so forth. This practice has also stopped. Now, the next available number is allocated to a new employee irrespective of role type or experience.

Meeting 4	13th February 2014 at the National Gallery of Ireland (NGI)
Attendees:	Leah Benson (LB) & Marian Reeves (MR)
Purpose:	Viewing Yeats archive

Along with the room dedicated to the Yeats archive, a large collection of Yeats sketchbooks are held in a different room and held (locked) in a specialised storage, to conserve the materials. We viewed some of the. LB then explained that these materials are protected by copyright: Jack B Yeats died in 1957 and the works are under copyright for 70 years after his death and given the volume of materials, it would cost around 850,000+ to purchase the rights to scan them and place on the web site, money which NGI does not have.

Also, next to this area are First Editions of books and plays by Jack B Yeats.

Anne Yeats, (Jack's niece, daughter of William Butler Yeats and) provided the materials to NGI. NGI had relatively little work to do with the materials (the role of the Archivists is to preserve the order of the archives) as Anne Yeats has organised and catalogued the materials and they had been preserved in good order. The meeting concluded.

Appendix D Meetings continued

Meeting 5	20th February 2014 at the National Gallery of Ireland (NGI)
Attendees:	Leah Benson (LB) & Marian Reeves (MR)
Purpose:	Entering Archive records and Searching + What systems are used

LB had mentioned (in a previous email) that Adlib was the most commonly used Archive software in Ireland. What systems are used at NGI? These are detailed below:

Doras (Irish word for door) is the system used to display the archives. This can be accessed via NGI website, from the Library and Archives page. See Thumbnail on the right hand side entitled **NGI Portico Doras.** Click on this and then click on the Doras image. (From the address bar in the browser we can see that the front-end is written in PHP).

Ibase: The back-end to Doras is **IBase,** a digital asset management system.

Yeats archive is not available online, except in documented format 11 Word documents, 1 for each collection.

Library / Bibliographic materials

Portico is the front end to the Bibliographic collection and Archives

Artwork Collections

eMuseum is used to display the collection. This is accessed via the NGI website. (From the Collection page select/click the thumbnail/button on the right hand side entitled 'online Collection'). eMuseum is provided by Gallery Systems and publishes / retrieves works from the NGI collection.

TMS (The Museum System) is the inventory control system a software package provided by Gallery systems and is used to manage the collection systems (TMS)

LB was due to present at the Virtual Cultural Heritage Ireland conference in Trinity, 27 and 28 March 2014. She talked about her presentation and the complexity of data in the archives. For example, one page in a scrap book might contain several completely different items: a letter, an image, a page with a sketch and description of the sketched object and so forth. Viewing this entry in the software shows that the archivist documents each of the individual items.

Appendix D Meetings continued

Meeting 6	05 May 2016 at the National Gallery of Ireland (NGI)
Attendees:	Raffaella Lanino, Documentation & Digitisation Assistant at NGI, Marian Reeves, PhD student, TCD
Purpose:	Understand Collection Management System (TMS) - purpose, who uses it, how collection is documented and any interconnections with other systems used by NGI

TMS is used to manage all objects in NGI permanent collections. The permanent collection includes paintings, sculpture, prints & drawings.

TMS is not used to record the Archives or Library materials and there is no connection to these systems. A system called Calm is used for the Archives. Still not sure what is used for Bib/Library.

All transactions / events of an object are documented in TMS: e.g. Location, Acquisition, Loans, Insurance, Transport, Exhibitions (Catalogues, Contracts, Loans, Venue etc), Conservation.

A nightly export schedule exports objects (marked as allowed for public display) from TMS to an online web based system eMuseum. eMuseum is integrated into the NGI website via the option 'online collections'

RL is responsible for object data input/update on TMS. Data comes from Curators (Acquisition sheet, catalogue descriptions, labels and or forth). Historical exhibition details are being input and this data is derived from an ongoing internship programme at NGI, where interns research 1) art books and 2) the dossier files, to discover exhibitions in which an object was presented.

Policy developed by curators at NGI is to browse ULAN (Unified List of Artist Names from Getty Research Institute) to ensure Artist preferred name is correct and for any other variations (which NGI may use in their publications, web etc.). Having said that, RL stated that Dr. A Waiboer uses RKD (Art History Institute of Netherlands) because these are the experts on artists in the Northern European collection, for which he is responsible.

Note on *Acquisition*: An acquisition form is completed by the curators (purchase/received date, source name/institution, contract, price, insurance, transport, and tombstone data-see below) and the object is assigned a temporary number. *Accessioning*: When the acquisition is approved by the board a permanent NGI code is allocated to it and tombstone data is input (if it has not already been input against the temporary number). The object now becomes part of the permanent collection.

Object Documentation: The minimum data stored in TMS about an object is classed as *tombstone* data. This comprises: Artist name, birth & death year, object title, object creation year, object dimensions, medium (e.g. oil on canvas), and object credit line (e.g. Donated by Person Name, Year Donated). The description of an object is not always stored on computer.

Getty vocabularies (AAT, ULAN and TGN) are available within TMS.

Many metadata elements besides the tombstone data elements are available within TMS. Priority of documentation is given to recent/new acquisitions but there is an on-going programme to expand documentation of the permanent collection.

Exhibition data may also be input. For example, location, title, period of exhibition.

Context data may also be input. For example, events at specified locations that may be relevant to the object can be input. TGN is available for location specification. RL will send me blank screen shot of Context screen.

Media data may also be input. This refers to the digital image, for example a photograph of a painting. Source and Copyright may be specified here.

Relationships: The term *Constituents* is used to refer to *people or institutions* associated in some way with or possibly referred to by an object. For example, the artist name, donor name, sitter, institution previously associated with the object.

Parent-Child relationships can be defined *between objects* in TMS. For example, An Album (parent) may consist of 10 prints (each print is a child of the album)

See Also relationships can be defined *between objects* in TMS. Perhaps this might be used where objects are about a similar theme or comparison or contrast to other object?

Note: Relationships cannot be defined between objects in TMS and those in other external in-house systems or with web based systems.

Display data V Index data: For example, Display date c1910 – 1980 can have different index data stored, i.e. Begin Date: 1905 and End Date: 1950.

NGI has drawn up a set of standards to use for such instances. For example, c (circa) dates will have index Begin Date set to -5 years and End Date set to + 5 years. Index values for Century and ? dates are also governed by NGI Standards.

Other systems: **Capture:** The Capture system embeds metadata into the digital image. This metadata is imported into Capture following its export from TMS. Capture is currently used by NGI for image capture, copy, pricing and sales.

Appendix E

7.1.6. Observation Notes and follow up questions

July 2013 and 5 August 2013

The following exhibitions are running at The National Gallery of Ireland:

- Masterpieces from the Collection
- The World Of Performance (commencing 10 August)
- From Galway to Leenane: Perceptions of Landscape
- Shades of Grey
- MDP (Master Development Plan of Restoration works at the gallery)

The exhibition visited is entitled **Masterpieces from the Collection** and is located in the Beit wing (rooms 1-10), and is one of 5 exhibitions currently in the National Gallery of Ireland (NGI).

The NGI describe **Masterpieces from the Collection** as containing some of the finest painting from its collection; including some from the European, early Renaissance through to twentieth century. Highlights of the Irish collection map the development of Irish art from the eighteenth century onwards and include key works by Hugh Douglas Hamilton, James Barry, William Orpen, John Lavery and Jack B. Yeats"

Jack B Yeats (1871 – 1957)

The Yeats display contains a selection of the artist's most popular works. Ten paintings are located in the Long Room and five are located in Room 2, in the Millenium wing.

1	For The Road (1951)
2	Men of Destiny (1946)
3	The Singing Horseman (1949)
4	In Memory of Boucicault & Bianconi (1937)
5	Above The Fair (1946)
6	Morning in the City (1946)
7	A Lake Regatta (1923)
8	Draughts (1922)
9	Before The Start (1915)
10	The Priest (1913)
11	Grief (1951)
12	Patriotic Airs (1923)
13	About to Write a Letter (1935)
14	Bachelors Walk, in Memory (1923)
15	Liffey Swim (1925)

The first 10 artworks are displayed along one wall in the Long room which has an entrance at either end. The paintings are arranged in the order listed above so depending on which entrance taken to the annex, the viewer encounters Yeats's latest paintings first and views the artworks back in almost chronological order to his earlier works or if using the alternate entrance the visitor will view the earlier paintings first up to the most recent.

Each painting is assigned a unique code / NGI No which is clearly displayed on the label. A label of plain white with black printing showing the following data:

- Artist Name,
- Place of Birth, Birth year – Year of Death)

- The Irish title of the work, (the year)
- The English title of the work, (the year)

- Presented by , Year (in Irish)
- Presented by , Year (in English)

- Provenance: e.g. Presented, Jack B. Yeats National Loan Exhibition Committee, 1945
- NGI Identifier e.g. 1769

Artworks 11 to 15 are displayed on two opposite walls. Grief dominates the right wall (Refer to Map), the adjacent wall is dominated by 'The Artists Studio (Lavery, 1909-1913) and the remaining artworks (2 to 4) are placed on the opposite wall. The paintings are arranged in the order listed above.

Dimensions: These paintings (1-5) above are smaller in size than the later works in the annex, except for Grief (1951) which is a large oil painting, Expressionist style?

The labels in room 2 are completely different than those for the painting in the Long room. The labels are cherry red and grey and show the following:
Each artwork is assigned a unique code / NGI No which is clearly displayed on the label.
They include a short description of the artwork which is helpful to the viewer.

Questions arising

1. What criteria were used to select these artworks for this exhibition?
2. What criteria were used in arranging them – not strictly chronological?
3. What criteria were used in deciding which artworks would be displayed in the main hall and in the annex?

Questions added following meeting 1

4. The data on the labels refers to Provenance e.g. on Loan, Bequeathed, Purchased 1965 (Shaw Fund) – What other types of transactions or events are there?
5. What data is captured when artworks are acquired or disposed of?
6. What procedures and specific tasks apply to acquisition, disposal, preservation or loan events and what data is recorded?
7. There is a description of the artwork on some labels, why not on all labels?
   Artwork Identifiers
8. Each artwork has a unique Identifier
9. How are NGI ID allocated – is it simply consecutive numbers as the Gallery receives in a new artwork?
10. How are identifiers generated for Artworks that are on loan? They do not have an NGI identifier and use a different numbering system e.g. artwork sample identifier is L.2009.1

Appendix E

7.1.7. Observations at the gallery

Exhibition: Masterpieces from the Collection - Irish Section

No.	NGI ID	Artist Forename	Artist Surname		Birth Year	Birth Location	To Year	Death location	Artwork		From Year	To Year		Year
Room 5														
1		Garreb	Morphy		?	Dublin		Dublin	Portrait William 4th Viscount Molyneaux	c.	1655	1717	c.	1700
2		Charles	Jervas	c.	1675	Offaly	1739	London	Portrait General William Cosby	c.	1690	1736		1710
3		Hugh	Douglas		1740	Dublin	1808	London	Frederick Harvey - Bishop of Derry				c.	1790
4		Jonathon	Fisher		1763	Active	1809	Dublin	A view of the Lower Lake, Killarney		1780's			
5		Nathaniel	Hone		1718	Dublin	1784	London	The Conjuror		1775			
6		Thomas	Roberts		1748	Waterford	1777	Lisbon	A Landscape		1770's			
7		James	Barry		1741	Cork	1806	London	Self Portrait as Timanthes	c.	1780	1803		
8		William	Ashford		1746	Birmingham	1824	Dublin	A View of Dublin from Chapelizod		1795	1798		
9		George	Barret		1728	Dublin	1784	London	A View of Powerscourt Waterfall	c.				1760
Room 4														
1		William	Davis		1812	Dublin	1873	London	A View of the rye Water near Leixlip		1850's			
2		Matthew James	Lawless		1837	Dublin	1864	London	A Sick Call					1863
3		James Arthur	O'connor		1792	Dublin	1841	London	A thunderstorm: The Frightened Wagoner					1832
4		Francis	Darby		1793	Waterford	1861	Exmouth	The Opening of the 6th Seal					1828
5		Joseph Patrick	Haverty		1794	Galway	1864	Dublin	The Blind Piper; Patrick O'Brien	c.				1844
6		Erskine	Nicol		1825	Leith	1904	Middlesex	The 16th, 17th (St. Patricks Day) & 18 March					1856

Appendix E Observations at the gallery

Exhibition: Masterpieces from the Collection - Irish Section

No.	NGI ID	Artist Forename	Artist Surname		Birth Year	Birth Location	To Year	Death location	Artwork		From Year	To Year		Year
Room 3														
1		Roderick	O'connor		1860	Roscommon	1940	Neuil-sur-Layon	La Jeune Bretonne	c.				1895
2		Harry	Jones Thaddeus		1860	Cork	1929	Ryde, Isle of Wight	The fishermans Mother	c.				1893
3		Helen Mabel	Trevor		1831	Loughbrickland, Co. Down	1900	Paris	Mass in Conemara Cabin					1883
4		Sir John	Lavery		1856	Belfast	1941	Rosennarra, co. kilkenny	Return from market					1884
5		Walter Frederick	Osborne		1859	Dublin	1903	Dublin	The Dublin Streets; a Vendor of Books					1889
6		William John	Leech		1881	Dublin	1968	Guilford	A Convent Garden, Brittany	c.				1913
Room 2														
1		Margaret	Clarke		1888	Newry	1961	Dublin	Ophelia	c.				1926
2		Gerard	Dillon		1916	Belfast	1971	Dublin	The little Green fields, Connemara					1945
3	1769	Jack	Butler Yeats		1871	London	1957	Dublin	Grief					1951
4		Mainie	Jellett		1897	Dublin	1944	Dublin	An Abstract Composition	c.	1930's			
5		Sir John	Lavery		1856	Belfast	1941	Rosennarra, co. kilkenny	The Artisis Studio		1909/ 1913			
6	2011.1	Jack	Butler Yeats		1871	London	1957	Dublin	Patriotic Airs					1923
7	1766	Jack	Butler Yeats		1871	London	1957	Dublin	About to write a Letter					1935
									Purchased 1964					
8		Paul	Henry		1876	Belfast	1958	Enniskerry	Connemara Village		1933	1934		
9	L.2009.1	Jack	Butler Yeats		1871	London	1957	Dublin	Batchelor's Walk, in Memory					1915
									On Loan from a private collection					
10	941	Jack	Butler Yeats		1871	London	1957	Dublin	The Liffery Swim					1923
									Presented, Havery Trust 1931					

Appendix E Observations at the gallery

Exhibition: Masterpieces from the Collection - Irish Section

No.	NGI ID	Artist Forename	Artist Surname		Birth Year	Birth Location	To Year	Death location	Artwork		From Year	To Year		Year
Long Room														
1	4309	Jack	Butler Yeats		1871	London	1957	Dublin	For The Road					1951
2	1134	Jack	Butler Yeats		1871	London	1957	Dublin	Men of Destiny					1946
3	4524	Jack	Butler Yeats		1871	London	1957	Dublin	The Singing Horseman					1949
4	4206	Jack	Butler Yeats		1871	London	1957	Dublin	In Memory of Boucicault & Bianconi					1937
5	1147	Jack	Butler Yeats		1871	London	1957	Dublin	Above the Fair					1946
6	1050	Jack	Butler Yeats		1871	London	1957	Dublin	Morning in the City					1937
7	1406	Jack	Butler Yeats		1871	London	1957	Dublin	A Lake Regatta					1923
8	1407	Jack	Butler Yeats		1871	London	1957	Dublin	Draughts					1922
9	1549	Jack	Butler Yeats		1871	London	1957	Dublin	Before the Start					1915
10	1804	Jack	Butler Yeats		1871	London	1957	Dublin	The Priest					1913

Appendix E Observations at the gallery

Exhibition: Masterpieces from the Collection - Irish Section

No.	NGI ID	Artist Forename	Artist Surname		Birth Year	Birth Location	To Year	Death location	Artwork		From Year	To Year		Year
Long Wall - Open Area														
1		James	Latham		1696	Tipperary	1747	Dublin	Bishop Robert Clayton & his wife Katherine	c.				1740
2		William	van Der Hagen		1720	Active	1745	?	Capriccio with Pastoral Figures					1732
3		Daniel	Maclise		1806	Cork	1870	London	A Scene from Gil Blas					1839
4		Richard	rothwell		1800	Athlone	1868	Rome	Calisto	c.	1850's			
5		Nathaniel	Hone		1831	Dublin	1917	Dublin	A View of Vulefranche from the East					1881
6		Richard Thomas	Moynan		1856	Dublin	1906	Dublin	Military Manouvres					1891
7		George	Collie		1904	Carrickmacross	1984	Dublin	The Midday Meal	c.				1927
8		Sean	Keating		1889	Limerick	1978	Dublin	An Allegory	c.				1922
9		Sir william	Orpen		1878	Stillorgan	1931	London	Count John McCormack (1884 - 1945) Tenor					1923
10		Sir John	Lavery		1856	Belfast	1941	Rosennarra, co. kilkenny	Portrait of Lady Lavery as Kathleen ní Houlihan					1927
11		William	Scott		1913	Greenock, Renfrewshire	1989	Coleford, Somerset	Frying Pan, Eggs & Napkins					1950
12		William	Crozier		1930	Glasgow	2011	Wickham, Hampshire	Flanders Fields		1961/ 1962			
13		Colin	Davidson		1968	Belfast			Michael Longley, celebrated Irish Poet		2011	2012		

Appendix F

7.1.8. Ethics

TRINITY COLLEGE DUBLIN

INFORMED CONSENTFORM

LEAD RESEARCHERS: This research is being conducted by Marian Reeves, a post-graduate student completing a PhD in Computer Science at Trinity College Dublin. It is being conducted under the supervision of Professor Khurshid Ahmad, Chair of Computer Science, Trinity College Dublin and in co-operation with domain experts at the National Gallery of Ireland (NGI).

BACKGROUND OF RESEARCH:

We are exploring the potential of semantic web technologies to provide an enhanced representation of fine art, specifically paintings, in support of research in fine art.

Goal of this research

The aim of this research is to incorporate the knowledge of museum curators whose work comprises the whole lifecycle of cultural heritage artefacts: from discovery to acquisition, to restoration and display. Curators deal with aesthetic objects such as paintings and a range of collectible objects from sketches, to correspondence, to critical reviews and to other artefacts held in the archives. By systematically organising these heterogeneous objects in an ontology and knowledge base the study examines how these structures may facilitate a more comprehensive representation of fine art and its relationships to collateral materials, than is typically available online.

PROCEDURES OF THIS STUDY:

Semi structured interview will be conducted with at least two domain experts involved in the curation of collections of aesthetic objects and archival materials at the gallery / museum or academic institution. The interviews will be for approximately 50 minutes and will be audio recorded. Interview transcripts will be provided to participants for review and approval.

An online survey will also be conducted. Participants in the survey will be art specialists. Survey duration is approximately 15 minutes.

PUBLICATION: The results of this research are required to be submitted for the PhD course at Trinity College and may be submitted for publication in academic conferences and journals. Individual results will be aggregated anonymously and research reported on aggregate results.

DECLARATION:

- I am 18 years or older and am competent to provide consent.
- I have read, or had read to me, a document providing information about this research and this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction and understand the description of the research that is being provided to me.
- I agree that my data is used for scientific purposes and I have no objection that my data is published in scientific publications in a way that does not reveal my identity.
- I understand that if I make illicit activities known, these will be reported to appropriate authorities.
- I understand that I may stop electronic recordings at any time, and that I may at any time, even subsequent to my participation have such recordings destroyed (except in situations such as above).
- I understand that, subject to the constraints above, no recordings will be replayed in any public forum or made available to any audience other than the current researchers/research team.
- I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights.
- I understand that I may refuse to answer any question and that I may withdraw at any time without penalty.
- I understand that my participation is fully anonymous and that no personal details about me will be recorded, *unless I explicitly consent to this*.

- *<If the research involves viewing materials via a computer monitor>* I understand that if I or anyone in my family has a history of epilepsy then I am proceeding at my own risk.
- I have received a copy of this agreement.

PARTICIPANT'S NAME: ……………………………………………………………………………

PARTICIPANT'S SIGNATURE: …………………………………………………………………

Date: ……………………………………………

Statement of investigator's responsibility: I have explained the nature and purpose of this research study, the procedures to be undertaken and any risks that may be involved. I have offered to answer any questions and fully answered such questions. I believe that the participant understands my explanation and has freely given informed consent.

RESEARCHERS CONTACT DETAILS:

Email: reevesm@tcd.ie…………………………………………………………………………

INVESTIGATOR'S SIGNATURE: ………………………………………………………………………

Date: ……………………………………………

SCSS Research Ethics Application Form August 2014

TRINITY COLLEGE DUBLIN

INFORMATION SHEET FOR PROSPECTIVE PARTICIPANTS

BACKGROUND OF RESEARCH:

We are exploring the potential of semantic web technologies to provide an enhanced representation of fine art, in support of research in fine art.

Goal of this research

The aim of this research is to incorporate the knowledge of museum curators whose work comprises the whole lifecycle of cultural heritage artefacts: from discovery to acquisition, to restoration and display. Curators deal with aesthetic objects such as paintings and a range of collectible objects from sketches, to correspondence, to critical reviews and to other artefacts held in the archives. By systematically organising these heterogeneous objects in an ontology and knowledge base the study examines how these structures may facilitate a more comprehensive representation of fine art and its relationships to collateral materials, than is typically available online.

Who is conducting this research? This research is being conducted by Marian Reeves, a post-graduate student completing a PhD in Computer Science at Trinity College Dublin. It is being conducted under the supervision of Professor Khurshid Ahmad, Chair of Computer Science at Trinity College Dublin and in co-operation with domain experts involved in the curation of collections at the National Gallery of Ireland.

There is no conflict of interest to declare. The research is not funded by any entity and is not carried out for commercial gain.

Your participation – in a semi structured interview

The purpose of the interview is to gain an understanding of the important functions you carry out as curator of the collections at the gallery / museum / institution. The interview will elucidate the processes you undertake and the different types of resources you consult. We wish to know what different kinds of information about a painting you believe to be crucial at different points in the curatorial life cycle; the sources of information about the objects; how and where information is recorded; the standards used for documentation of objects and your opinion of the value that collateral materials can add to knowledge about an artwork and to the scholarly research of artworks.

The interview will take about 50 minutes to complete. Your participation is greatly appreciated!

Participation criteria: Participants must be at least 18 years of age. This interview is restricted to art specialists responsible for the curation of collections at a gallery or museum or academic institution.

Agreement

- Participation in the study is entirely voluntary. You have the right to opt in and subsequently to withdraw at any time for any reason without penalty.
- All participants must read and sign the Consent Form before the interview takes place.
- If illicit activities are made known then they will be reported to the appropriate authorities.
- Your identity will not be revealed in any way. You will remain anonymous in the analysis and subsequent publication and presentation of the resulting data and findings.
- You and/or your institution may be identified in subsequent research publication and presentation *only if you expressly consent* to either one or both of these pieces of information being made public.
- The interview will be audio recorded. A transcript of your interview will be provided to you for review and approval. The audio recording will be made available to the researcher and supervisor only and will not be replayed in any public forum or presentation of the research.
- You have the option of omitting questions you do not wish to answer.
- You can have the recording stopped at anytime during the interview if you so choose and have the interview destroyed.
- There are no anticipated risks to your involvement in this study

TRINITY COLLEGE DUBLIN

INFORMATION SHEET FOR PROSPECTIVE PARTICIPANTS

We are exploring the potential of semantic web technologies to provide an enhanced representation of fine art, specifically paintings, in support of research in fine art.

Goal of this research

The aim of this research is to incorporate the knowledge of museum curators whose work comprises the whole lifecycle of cultural heritage artefacts: from discovery to acquisition, to restoration and display. Curators deal with aesthetic objects such as paintings and a range of collectible objects from sketches, to correspondence, to critical reviews and to other artefacts held in the archives. By systematically organising these heterogeneous objects in an ontology and knowledge base the study examines how these structures may facilitate a more comprehensive representation of fine art and its relationships to collateral materials, than is typically available online.

Who is conducting this research?

This research is being conducted by Marian Reeves, a post-graduate student completing a PhD in Computer Science at Trinity College Dublin. It is being conducted under the supervision of Professor Khurshid Ahmad, Chair of Computer Science at Trinity College Dublin and in co-operation with domain experts involved in the curation of collections at the National Gallery of Ireland.

There is no conflict of interest to declare. The research is not funded by any entity and is not carried out for commercial gain.

Your participation – in a survey

The survey will help us to understand the processes you undertake and the information about an artwork that you consider important for scholarly research of an artwork. In particular, we wish to know what types of resources you consult, the different kinds of information you believe are important to enhance knowledge of the artwork under study; your opinion of the value that collateral materials such as the artist sketches, correspondence and other artist memorabilia might contribute to your research and your understanding of and knowledge about an artwork.

The survey takes about 15 minutes to complete. URL: http://www.bvl.ie/artsurvey1/index.php

Participation criteria: Participants must be at least 18 years of age. The survey is restricted to art specialists, such as, those who are responsible for the curation of collections at a gallery or museum or academic institution; those who organise art exhibitions; conduct research of artworks as part of their teaching, as part of their studies or for academic research for conference or journal publication.

Agreement

- Participation in the study is entirely voluntary. You have the right to opt in and subsequently to withdraw at any time for any reason without penalty.
- All participants must read and sign/tick the online Consent Form before commencing the survey.
- You have the option of omitting questions you do not wish to answer.
- Your identity will not be revealed in any way. You will remain anonymous in the analysis and subsequent publication and presentation of the resulting data and findings.
- You and/or your institution may be identified in subsequent research publication and presentation *only if you expressly consent* to either one or both of these pieces of information being made public.
- There are no anticipated risks to your involvement in this study.
- Please do not name third parties in any open text field of the questionnaire. Any such replies will be anonymised.
- In the extremely unlikely event that illicit activity is reported I will be obliged to report it to appropriate authorities.

Appendix G

7.1.9. Sample SQL Scripts

Script to create the agent / artist table

```
CREATE TABLE a_agent(
agentID INT UNSIGNED NOT NULL AUTO_INCREMENT,
agenttypeID varchar(20) default "Artist",
agentdisplayname varchar(50),
agentIndexName varchar(50) NOT NULL,
gender VARCHAR(1) NOT NULL DEFAULT "M",
namesourcebiblio integer not NULL default 1,
displaybiography VARCHAR(1000) NOT null DEFAULT "Here is a biography of the artist, person or
corporate body",
BirthDateDisplay varchar(20) not NULL default "19th Century",
DeathDateDisplay varchar(20),
fdateflg VARCHAR(1) DEFAULT "X",
tdateflg VARCHAR(1) DEFAULT "X",
earliestdate INT,
latestdate INT,
BirthPlaceDisplay varchar(50),
DeathPlaceDisplay varchar(20),
nationID varchar(10),
agentNote varchar(1000),
refID varchar(20) default "500002320",
refvocab varchar(20) default "Getty ULAN",
vocaburi varchar(150) default
"http://www.getty.edu/vow/ULANFullDisplay?find=Yeats&role=&nation=&prev_page=1&subjectid=500
002320",
listnum SMALLINT UNSIGNED DEFAULT 1,
updatedate TIMESTAMP NOT NULL,
updateuserid VARCHAR(25) NOT NULL,
PRIMARY KEY  (agentID),
FOREIGN KEY (nationID) references a_nation(nationID)
)ENGINE=InnoDB DEFAULT CHARSET=utf8;
```

Sample script to delete all persons and then populate the table with two persons / artists

```
DELETE FROM a_agent;
INSERT INTO `a_agent (agentID, agenttypeID, agentdisplayname, agentIndexName, gender,
namesourcebiblio, displaybiography, BirthDateDisplay, DeathDateDisplay, fdateflg, tdateflg,
earliestdate, latestdate, BirthPlaceDisplay, DeathPlaceDisplay, nationID, agentNote, refID, refvocab,
vocaburi, listnum, updatedate, updateuserid) VALUES ('1', 'ARTIST', 'Jack B Yeats', 'Yeats,Jack B',
'M', 1,'Here is a biography of the artist, person or corporate body', '29 August 1871', '28 March 1957',
'X', 'X', '1871', '1957', 'London', 'Dublin', 'IE', NULL, '500002320', 'Getty ULAN',
'http://www.getty.edu/vow/ULANFullDisplay?find=Yeats&role=&nation=&prev_page=1&subjectid=500
002320', '1', CURRENT_TIMESTAMP, '');

INSERT INTO a_agent (agentID, agenttypeID, agentdisplayname, agentIndexName, gender,
namesourcebiblio, displaybiography, BirthDateDisplay, DeathDateDisplay, fdateflg, tdateflg,
earliestdate, latestdate, BirthPlaceDisplay, DeathPlaceDisplay, nationID, agentNote, refID, refvocab,
vocaburi, listnum, updatedate, updateuserid) VALUES ('2', 'ARTIST', 'Johannes Vermeer',
'Vermeer,Johannes', 'M', 1,'Here is a biography of the artist, person or corporate body', 'Baptized 31
Oct 1632', '15 December 1675', 'X', 'X', '1871', '1957', 'Delft', 'Delft', 'NL', NULL, '500005593', 'Getty
ULAN',
'http://www.getty.edu/vow/ULANFullDisplay?find=Vermeer&role=&nation=&prev_page=1&subjectid=5
00005593', '1', CURRENT_TIMESTAMP, '');
```

Sample script to create a View that links an individual archival objects (e.g. a sketch/ photograph) to the associated archival description (ISAD(G) record/tuple

```
CREATE VIEW vwarchobjAllDetails1 AS
SELECT t_archobject.archobjectID,
t_archobject.eadid,
t_archobject.archobjectname,
t_archobject.archimgurl,
t_archobject.archobjecttypeID,
t_archobject.archobjectnotes,
t_archobject.archobjecttranscription,
t_archobject.mimetypeID,
t_archobject.tqmediumtypeID,
t_objecttype.objecttypename,
t_mimetype.mimetypedesc,
t_archmain.archmainref,
t_archmain.unittitle,
t_archmain.unitdate,
t_archmain.scopecontent,
t_tqmediumtype.tqmediumtypename
FROM t_archobject
LEFT JOIN t_objecttype ON (t_archobject.archobjecttypeID = t_objecttype.objecttypeID)
LEFT JOIN t_tqmediumtype ON (t_archobject.tqmediumtypeID = t_tqmediumtype.tqmediumtypeID)
LEFT JOIN t_mimetype ON (t_archobject.mimetypeID = t_mimetype.mimetypeID)
LEFT JOIN t_archmain ON (t_archobject.eadid = t_archmain.eadid);
```

Appendix H

7.1.10. Getty Vocabularies

7.1.10.1. Sample ULAN record from Getty Research Institute

Source: http://www.getty.edu/research/tools/vocabularies/ulan/about.html

Date Obtained: 6 April 2018

The **Union List of Artist Names** ® (ULAN), the Art & Architecture Thesaurus ® (AAT), the Getty Thesaurus of Geographic Names ® (TGN). the Cultural Objects Name Authority ® (CONA), and the Iconography Authority (IA) are structured resources that can be used to improve access to information about art, architecture, and material culture. Through rich metadata and links, it is hoped that the Getty vocabularies will provide a powerful conduit for research and discovery for digital art history and related disciplines.

Cataloging: They may be used as data value standards at the point of documentation or cataloging. In this context, they may be used as a controlled vocabulary or authority by the cataloger or indexer; they provide preferred names/terms and synonyms for people, places, and things. They also provide structure and classification schemes that can aid in documentation.

Retrieval: They may be used as search assistants in database retrieval systems and broader linked data applications. They are knowledge bases that include semantic networks that show links and paths between artists; these relationships can make retrieval more successful.

Research tools: They may be utilized as research tools, valuable because of the rich information and contextual knowledge that they contain.

Sample Record for Jack B Yeats

ID: 500002320
Record Type: Person
Page Link: http://vocab.getty.edu/page/ulan/500002320
Yeats, Jack B. (Irish painter, 1871-1957)
Note: He was the brother of the poet W. B. Yeats and the son of artist Jack B. Yeats. For the first 20 years of his career Yeats produced illustrations and genre and landscape paintings. In the late 1880s he began to be known as a portrait painter and talented draftsman. Comment on works: Landscapes

Yeats, Jack B. (**preferred**,V,index)
Jack B. Yeats (V,display)
Yeats, Jack Butler (V,LC)
Jack Butler Yeats (V)
Yeats, John Butler (U)
Yeats (LU,U)
Nationalities: Irish (**preferred**) British (modern)
Roles: artist (**preferred**), painter
Gender: male
Birth and Death Places:
Born: Tullylish (Armagh, Banbridge and Craigavon, United Kingdom) (parish (ecclesiastical))
Died: New York City (New York state, United States) (inhabited place)
Related People or Corporate Bodies:
child of .... Yeats, John Butler (Irish painter, 1839-1922) [500016889)
sibling of.... Yeats, William Butler , (Irish poet and painter, 1865-1939) [500019359]

List / Hierarchical Position:
Persons, Artists
Yeats, Jack B. (I)
Biographies:
(Irish painter, 1871-1957) ..... [BHA Preferred]
(Irish painter 1871-1957) ..... [GRLPSC]
(Irish painter, 1871-1957) ..... [PROV]
(British artist, 1871-1957) ..... [WCP]
(British artist, 1871-1957) ..... [WCI]
(British artist, 1871-1957) ..... [WL-Courtauld]
(artist, 1871-1957) ..... [GRL]
(artist, active 20th century) ..... [GRISC]
(Irish artist, 1871-1957) ..... [PROV]

Additional Names:

Sources and Contributors:
Jack Butler Yeats [PROV, VP], Getty Provenance Index Databases [online] (1999), Provenance Index Databases, Authority file (1985-)
Jack B. Yeats [GRISC, VP] Getty Vocabulary Program rules
Harald Szeemann papers, Finding Aid, GRI Special Collections (2011)
Yeats .... [PROV]
Provenance Index Databases, Authority file (1985-)
Yeats, Jack B. [GRISC Preferred, VP Preferred]
Grove Dictionary of Art online (1999-2002) "Jack B. Yeats." Accessed 03/04/04
Harald Szeemann papers, Finding Aid, GRI Special Collections (2011-)

Subject: ....

[BHA, Gallery Systems, GRISC, GRL, GRLPSC, PROV, VP, WCI, WCP, WL-Courtauld]
Bénézit, Dictionnaire des Peintres (1976)
Gallery Systems (2000-) LOC ID: n50020999
Getty Provenance Index Databases [online] (1999)
Grove Dictionary of Art online (1999-2002) "Jack B. Yeats." Accessed 03/04/04
Harald Szeemann papers, Finding Aid, GRI Special Collections (2011-)
Houfe, Dictionary of British Book Illustrators (1978)
Library of Congress Authorities database (n.d.) LOC ID: n50020999
McGraw-Hill Dictionary of Art (1969)
RILA/BHA (1975-2000)
Vollmer, Künstler-Lexikon 20. Jahrhunderts (1953-62)
Witt Library, Authority files
Note: English
[PROV, VP, WL-Courtauld]
Grove Dictionary of Art online (1999-2002) Accessed 07/24/2002]
Provenance Index Databases, Authority file (1985-)

Appendix H Vocabularies continued

7.1.10.2. Sample AAT record from Getty Research Institute

Source:

http://www.getty.edu/vow/AATFullDisplay?find=baroque&logic=AND&note=&subjectid=300021147

Date Obtained: 6 April 2018

Record Type: concept
ID: 300021502
Page Link: http://vocab.getty.edu/page/aat/300021502

Expressionist (style) (<modern European styles and movements>, European, ... Styles and Periods (hierarchy name))
Note: Refers to an international style of art, literature, music, dance, and theater that flourished between 1905 and 1920, especially in Germany. The style is characterized by the abandonment of traditional standards of realism and proportion in favor of expressing the artist's emotions, resulting in distortions of line, color, and form.

Terms:
Expressionist (style) (**preferred**,C,U,English-P,D,U,A)
Expressionism (C,U,LC,English,AD,U,U)
表現主義 (C,U,Chinese (traditional)-P,D,U,U)
biǎo xiàn zhǔ yì (C,U,Chinese (transliterated Hanyu Pinyin)-P,UF,U,U)
biao xian zhu yi (C,U,Chinese (transliterated Pinyin without tones)-P,UF,U,U)
piao hsien chu i (C,U,Chinese (transliterated Wade-Giles)-P,UF,U,U)
expressionistisch (C,U,Dutch-P,D,U,U)
Expressionisme (C,U,Dutch,AD,U,U)
Expresionismo (C,U,Spanish-P,D,U,U)

Facet/Hierarchy Code: F.FL

Hierarchical Position:
Styles and Periods Facet
.... Styles and Periods (hierarchy name) (G)
.... <styles, periods, and cultures by region> (G)
.... European (G)
.... <modern European styles and movements> (G)
.... Expressionist (style) (G)

Additional Notes:
Chinese (traditional) ..... 在1905至1920年間蓬勃發展的國際風格，包括藝術、文學、音樂、和劇場，在德國尤其盛行。此風格的特色為捨棄傳統的寫實主義和比例標準，偏向表現藝術家的情緒，造成線條、色彩和形狀的扭曲。
Dutch ..... Verwijst naar een internationale stijl van kunst, literatuur, muziek, dans en theater die, vooral in Duitsland, floreerde tussen 1905 en 1920. De stijl wordt gekenmerkt door het opgeven van de traditionele normen van realisme en proportie in het voordeel van de uiting van de emoties van de kunstenaar resulterend in verstoringen van de lijn, kleur en vorm. Periode 1910-1930.
Spanish ..... Se refiere a un estilo internacional de arte, literatura, música, danza y teatro que floreció entre los años 1905 y 1920, especialmente en Alemenia. El estilo es caracterizado por el abandono de las normas tradicionales de realismo y proporción a favor de la expresión de las emociones del artista, causando las distorsiones de línea, color, y la forma.

Related concepts:
followed .... **Proto-Expressionist**
.... (<modern European styles and movements>, European, ... Styles and Periods (hierarchy name)) [300403931]
has associates .... **Expressionists (artists)**
.... (<artists by subject or style of work>, artists (visual artists), ... People (hierarchy name)) [300389788]

Sources and Contributors:

表現主義............ [AS-Academia Sinica Preferred]
.... 台灣古建築圖解事典 178-179
.... 國家教育研究院－雙語詞彙、學術名詞暨辭書資訊網 28 July, 2014 August 26, 2011
.... 大英百科全書線上繁體中文版 August 26, 2011
.... 大英視覺藝術百科全書 Vol. 1, 178
.... 智慧藏百科全書網 August 26, 2011
.... 朗文當代大辭典 598
.... 牛津當代大辭典 614
.... 藝術名詞與技法辭典 165
.... 雄獅西洋美術辭典 969

biao xian zhu yi............ [AS-Academia Sinica]
.... AS-Academia Sinica

biǎo xiàn zhǔ yì............ [AS-Academia Sinica]
.... AS-Academia Sinica

Expresionismo............ [CDBP-DIBAM Preferred]
.... Fatás y Borrás, Diccionario de términos de Arte y elementos de Arqueología, Heráldica y Numismática (1997) 107

Expressionism............ [VP]
.... Avery Index (1963-) Expressionism (Architecture)
.... CDMARC Subjects: LCSH (1988-) Expressionism (Art)
.... RIBA, Architectural Keywords (1982) Expressionism (Architecture)
.... RILA, Subject Headings (1975-1990)
.... Worcester Art Museum Library, List of subject headings, unpub. (1976)

Expressionisme............ [RKD, AAT-Ned]
.... Van Dale groot woordenboek (1994)

Expressionist (style)............ [GCI Preferred, VP Preferred]
.... AATA database (2002-) 123504 checked 26 January 2012
.... Arnason, History of Modern Art (1977) 178
.... Mayer, Dictionary of Art Terms and Techniques (1969) 139
.... Watkin, History of Western Architecture (1986) 513

expressionistisch............ [RKD, AAT-Ned Preferred]
.... Van Dale groot woordenboek (1994)

piao hsien chu i............ [AS-Academia Sinica]
.... AS-Academia Sinica

Subject: ..... [AS-Academia Sinica, CDBP-DIBAM, GCI, RKD, AAT-Ned, VP]
.... AATA database (2002-) 123504 checked 26 January 2012
.... Van Dale groot woordenboek (1994)

Note:

English .... [VP]
.... Legacy AAT data

Chinese (traditional).......... [AS-Academia Sinica]
.... AS-Academia Sinica

Dutch .... [RKD, AAT-Ned]
.... AAT-Ned (1994-)
.... Van Dale groot woordenboek (1994)

Spanish .... [CDBP-DIBAM]
.... TAA database (2000-)

Appendix H Vocabularies continued

7.1.10.3. Sample CONA record from Getty Research Institute

Source: http://www.getty.edu/research/tools/vocabularies/cona/about.html#sample

Date Obtained: 6 April 2018

The **Cultural Objects Name Authority** ® (CONA) compiles titles, attributions, depicted subjects, and other metadata about works of art, architecture, and cultural heritage, both extant and historical, physical and conceptual. Metadata is gathered or linked from museum collections, special collections, archives, libraries, scholarly research, and other sources. CONA is linked to the Getty's structured vocabularies, the Art & Architecture Thesaurus ® (AAT), the Getty Thesaurus of Geographic Names ® (TGN), and the Union List of Artist Names ® (ULAN). Through rich metadata and links, CONA may provide a powerful conduit for research and discovery for digital art history and related disciplines.

Subject_ID: 700000299

Record Type: movable work
Catalog Level: item
Object / Work Type: etching (print)
Classification: prints

TITLES
Title / Name: Dessein de l'illumination et du feu d'artifice donné a Monseigneur le Dauphin a Meudon le 3e septembre 1735
Preference: preferred
Type: repository
Language: French
Contributor: GRISC
Source: GRI Special Collections Finding Aids databases (2012-) **Page:** 672066

Title / Name: Fireworks display at Meudon honoring Dauphin, Louis in 1735
Preference: variant/alternate
Type: descriptive
Language: English
Contributor: VP
Source: GRI Special Collections Finding Aids databases (2012-)

CREATOR
Creator Display: Cochin, Charles Nicolas (printmaker, 1715-1790), after Debonneval, M.
Related Person / Corp.Body Link: Cochin, Charles-Nicolas, the elder (French engraver, 1688-1754) [500008603]
Role: printmaker
Related Person / Corp.Body Link: Debonneval, M. (French royal official, active 1735) [500353675]
Role: designer

DATE
Creation Date: published 1735?
Start Date: 1735 **End Date:** 1740

LOCATIONS
Flag: current **Type:** N/A
Location: Getty Research Institute, Special Collections (Los Angeles, Los Angeles county, California, United States) [500353615]
Repository Numbers: (CMalG)1394-070; accession number: P950001** (FF. 28)
Credit Line: copyright J. Paul Getty Trust, Getty Research Institute, Special Collections
Address Note: 1200 Getty Center Dr., Suite 1100, Los Angeles, CA 90049-1688

Location Flag: Other **Type:** publication
Location: Paris (France)

STYLE / CULTURE
Culture: French

PHYSICAL CHARACTERISTICS
Material and Technique: 1 print: engraving, etching, black and white
Term: ink **Role:** medium **Flag:** material
Term: paper **Role:** support **Flag:** material
Term: engraving (printing process) **Role:** technique **Flag:** technique/implement
Term: etching (printing process) **Role:** technique **Flag:** technique/implement

Dimensions Display: plate mark 41.3 x 49.7 cm
Value: 41.3 **Unit:** cm **Type:** height **Extent:** platemark
Value: 49.7 **Unit:** cm **Type:** width **Extent:** platemark

Inscription: Debonneval invenit; C.N. Cochin filius delineavit et sculpsit . "Cette feste ordonnée par M. le duc de Gesvres pair de France, premier gentilhomme de la chambre du roy, à esté conduite par M. Debonneval intendant et controlleur de l'Argenterie menua plaisirs et affaires de la chambre de Sa Majesté ."--lower margin. Title engraved below the image on both sides of the Bourbon coat of arms.

Provenance: Jacques Vellekoop; London; Purchase; From the collection of Alan Brock and his son, Alan St. Hill Brock, English pyrotechnists

DEPICTED SUBJECT
General Subject:
Term: events **Pref:** preferred
Term: history and legend

Specific Subjects:
Term: ephemeral structure *[AAT 300007997]*
Term: Louis, Dauphin of France (French noble, 1729-1765) *[ULAN 500353671]*
Extent: honoree
Term: customs (social) *[AAT 300055807]*
Term: festivals *[AAT 300073472]*
Term: birthdays *[AAT 300069093]*
Term: Meudon (Île-de-France, France) *[TGN 7009340]*
Term: theater (discipline) *[AAT 300054148]*
Term: fireworks (visual works) *[AAT 300249836]*

NOTE
Descriptive Note:
The print depicts the fireworks display held September 3, 1735, at Meudon in celebration of the sixth birthday of the Dauphin, Louis; Against an ephemeral backdrop consisting of a colonnade and flaming structures, Hercules's twelfth labor is enacted before spectators; Forms part of Brock fireworks collection (Special Collections accn. no. P950001**); Trimmed close to plate mark.

Contributor: GRISC
Source: GRI Special Collections Finding Aids databases (2012-)

Appendix H Vocabularies continued

7.1.10.4. Sample IA record from Getty Research Institute

Source: http://www.getty.edu/research/tools/vocabularies/cona/about.html#sample_ia

Date Obtained: 6 April 2018

Purpose of IA

The Getty Iconography Authority (IA) includes proper names and other information for named events, themes and narratives from religion/mythology, legendary and fictional characters, themes from literature, works of literature and performing arts, and legendary and fictional places. Examples include events (e.g., ***American Civil War (United States History)***), mythological characters (e.g., ***Venus (Roman deity)***), legendary places (e.g.,***Xibalba (Mayan underworld)***), iconographic narratives and themes (e.g., ***Adoration of the Magi (New Testament narrative)***), and literary themes (e.g., ***The Inferno (poem, Dante, 1308/1321)***).

The IA includes subjects not described by AAT (generic terms), TGN (geographic names), ULAN (names of people and corporate body), or CONA works.

The IA is linked to the Getty vocabularies. The IA is linked to other iconography sources, including Library of Congress authorities and Iconclass.

The IA has a thesaural structure. It includes equivalence, associative, and hierarchical relationships.

History of IA

The Getty Iconography Authority (IA) is based on the Subject Authority of the ***Categories for the Description of Works of Art*** (CDWA).

To place the IA in context: CONA may be seen as an implementation of CDWA; the AAT, TGN, and ULAN correspond to the "authorities" described in CDWA. However, there was no Getty vocabulary for the fourth CDWA authority, which is an authority in which to record named iconographical subjects depicted in works. Nor was there any other source of linkable iconographical information that is global in scope and meeting the criteria of the CDWA Subject Authority thesaurus. Thus, in the creation of CONA, the Iconography Authority was implemented as a module within CONA, to record iconographical topics needed to index CONA records, and to link to Iconclass and other outside sources of iconography. In time, with contributions from the user community, the IA may become a separate resource, useful and used outside the context of CONA.

Sample IA Record

Iconography ID: 1000851
Iconography Parent: 1000959 (Buddhist characters)
Iconography Type: Character/Person

Descriptive Note: As a subject in art and literature, the primary protagonist in narratives surrounding the life, deeds, and teachings of Buddhism.

Display Date: from 6th century BCE
Start Date: -520 **End Date**: 9999

Label: Buddha (Buddhist characters, Buddhist iconography, Legend, Religion, Mythology)

NAMES:
(sequence, preferred/variant flag, name, language and preference, term type)

1 P Buddha (primary character of Buddhism) *(English-P, descriptor)*

Sources:
Agnew, Reed, Ball, Cave Temples of Dunhuang (2016)
Agnew, Conservation of Ancient Sites on the Silk Road (2010)
Pal, Indian Sculpture (1986-1988)
Ling, Dictionary of Buddhism (1972)
Buswell and Lopez, Princeton Dictionary of Buddhism (2013)

2 V Gautama *(English, UF)*

Source:
Encyclopedia Britannica Online (2002-)
3 V Gotama *(English, UF)*

Source:
Encyclopedia Britannica Online (2002-)
4 V Siddhartha (English, UF)

Source:
Ling, Dictionary of Buddhism (1972)
5 V Siddhatta (English, UF)

Source:
Ling, Dictionary of Buddhism (1972)
6 V Shakyamuni (English, UF)

Source:
Ling, Dictionary of Buddhism (1972)
7 V 佛 (Chinese-P, Descriptor)

Source:
Dunhuang Research Academy, Mogao Grottoes (2000-)

Associative Relationships (IA to IA)

Type: 5310 predecessor of **Name:** Maitreya (Buddhist bodhisattva)
Type: 5510 relative of **Name:** Ananda (Buddhist character)
Type: 5003 associated with **Name:** Kasyapa (Buddhist character)
Type: 5503 is protagonist for **Name:** Sakyamuni's Life of Austerities
Type: 5503 is protagonist for **Name:** Buddha in Jetavana preaching

AAT Relationships (IA to AAT)

Type: culture/religion is **Name:** Buddhism (religions, ...Associated Concepts Facet)

TGN Relationships (IA to TGN)
Type: flourished/active in **Name:** India (Asia, World)

ULAN Relationships (IA to ULAN)
Type: identified as **Name:** Buddha (Asian ascetic, founder of Buddhism, ca. 560-ca. 480 BCE)

Record Sources/Links:
Source: Iconclass 2100 Browser (2009-)
Page/Link: 12H13(BUDDHA)
Source: Library of Congress Authorities online (2002-)
Page/Link: sj 96004936

Appendix I

7.1.11. Metadata – ISAD(G)

General International Standard Archival Description – ISAD(G) comprises 26 elements.

Descriptive Elements	All 26 elements covered by these general rules are available for use, but only a subset is needed in any given description.
1 Reference Code 2 Title 3 Creation Dates 4 Level of Description 5 Extent and Medium 6 Creator(s) 7 Administrative/Biographical History 8 Archival History 9 Immediate Source of Acquisition 10 Scope and Content 11 Appraisal and Destruction 12 Accruals 13 Arrangement 14 Access Conditions 15 User Conditions 16 Language/Scripts of Material 17 Physical Characteristics and Technical Requirements 18 Finding Aids 19 Originals Information 20 Copies Information 21 Related Material 22 Publication Note 23 Note Control Elements 24 Archivist's Note 25 Rules/Conventions 26 Date of Description(s)	A very few elements are *considered essential for international exchange of descriptive information*: a. Reference code; b. Title; c. Creator; d. Date(s); e. Extent of the unit of description; and f. Level of description.

Examples of ISAD(G) records – at Fonds, Series and Item level for: **Corporate fonds (organizational records): description of the fonds level and one of its series and items. Language of description: English (U.S.A.)**

Fonds level

3.1.1 Reference code(s) US DNA 220

3.1.2 Title Records of the Commission on the Bicentennial of the United States Constitution

3.1.3 Date(s) 1983-1992 (dates of accumulation)

3.1.4 Level of description Fonds

3.1.5 Extent and medium of the unit of description (quantity, bulk, or size) 202 linear feet of textual records on paper, graphic materials on paper, photographic prints, and video recordings

3.2.1 Name of creator(s) Commission on the Bicentennial of the United States Constitution

3.2.2 Administrative history

The Commission on the Bicentennial of the United States Constitution was established by an act of Congress approved September 29, 1983, and extended until June 30, 1992. The Commission was composed of 23 members, with former Chief

Justice of the United States Warren E. Burger serving as its Chairman. The Commission's primary statutory obligation was to encourage private organizations and state and local governments to organize and articipate in bicentennial activities that commemorated the drafting, ratification, and history of the Constitution. The Commission sought cooperation, advice, and assistance from both private and governmental agencies and organizations and also delegated authority to State advisory commissions to assist in its efforts. In addition, the Commission served as a clearinghouse for the collection and dissemination of information about bicentennial events and plans. The chief focus of the Commission's programs was educational, with particular emphasis on primary and secondary schools. Accordingly, the Commission committed the major part of its resources to teaching materials, teacher training, and the development and dissemination of new publications on the Constitution.

3.4.3 Language / scripts of material English

Series level

3.1.1 Reference code(s) DNA NWDNS-220-BCC

3.1.2 Title Photographs of Nationwide Celebrations of the Bicentennial of the Constitution

3.1.3 Date(s) 1987 (dates of accumulation)

3.1.4 Level of description Series

3.1.5 Extent and medium of the unit of description (quantity, bulk, or size) 1700 photographic prints

3.3.1 Scope and content The photographs in this series consist primarily of coverage of state and local observances of Constitution Day, Sept. 17, 1987. There are also other celebrations represented, such as Independence Day. The photographs, which were submitted by local liaison offices to the Commission, show parades, elementary school displays and pageants, ceremonial tree plantings, and other activities. In addition, there are photographs documenting citizenship ceremonies, and showing new Americans taking the oath of citizenship. Nearly all of the photographs are captioned. Persons of interest photographed in this series include Arkansas Governor Bill Clinton, speaking at Ft. Smith's Constitution celebration, and former astronaut Edwin "Buzz" Aldrin at El Dorado, California's celebration.

3.3.2 System of arrangement Photographs are arranged alphabetically by state, thereunder alphabetically by city, town, or county.

3.4.1 Conditions governing access: Unrestricted.

3.4.2 Conditions governing reproduction: Photographs from commercial sources may be subject to copyright restrictions.

3.4.5 Finding aids: Container and folder list.

Item level

3.1.1 Reference code(s) US DNA NWDNS-220-BCC-1

3.1.2 Title America on Parade: America's Parade of History Sets Sail with Christopher Columbus and Soars Out into Space Aboard the Shuttle "Columbia"

3.1.3 Date(s) 1987 (date of creation)

3.1.4 Level of description Item

3.1.5 Extent and medium of the unit of description (quantity, bulk, or size) 1 poster on paper

3.3.1 Scope and content This poster is a time chart of events in United States history from 1492 to the 1980s.

3.4.2 Conditions governing reproduction May not be reproduced without the written consent of the National Geographic Society.

Source http://www.icacds.org.uk/eng/ISAD(G).pdf

7.1.12. Metadata - Sample VRA

ELEMENTS

No	Element Name	Sub elements	No. of discrete elements/sub elements
1	**work, collection, or image** *(id)*		1
2	**Agent**	Attribution	2
		Culture	3
		Dates (*type*)	4
		earliestDate *(circa)*	5
		latestDate *(circa)*	6
		Name (*type*)	7
		Role	8
3	**culturalContext**		9
4	**Date**	Date (*type*)	10
		earliestDate *(circa)*	11
		latestDate *(circa)*	12
5	**Description**		13
6	**Inscription**	Author	14
		Position	15
		Text (*type*)	16
7	**Location**	Location *(type)*	17
		Name (*type*)	18
		Refid (*type*)	19
8	**Material**	Material (*type*)	20
9	**Measurements**	Measurements (*type,unit)*	21
10	**Relation**	Relation (*type, relids*)	22
11	**Rights**	rights *(type)*	23
		rightsHolder	24
		Text	25
12	**Source**	Name (*type*)	26
		Refid (*type*)	27
13	**stateEdition**	stateEdition (*count,num,type)*	28
		description	29
		name	30
14	**stylePeriod**		31
15	**Subject**	Term (type)	32
16	**Technique**	technique	33
17	**Textref**	Name (*type*)	34
		Refid (*type*)	35
18	**Title**		36
19	**Worktype**		37

Original Source: http://www.loc.gov/standards/vracore/VRA_Core4_Outline.pdf

VRA Core 4.0 also includes attributes considered global or "floating" that can be used as needed to modify any element or sub-element rather than being tied to any specific one. For example:

Attribute	Description
extent	refers to the part of the work, image or collection being described by the element or sub-element that it modifies.
dataDate	refers to the date and/or time a particular piece of data was captured
href	refers to a hypertext reference that provides a link to another electronic resource
pref	indicates that a particular data value is the preferred value when multiple data values for the same element or sub-element exist
refid	refers to id numbers or codes coming from the local institution or resource named in the source attribute
rules	refers to any data content standards used to construct the value recorded in the element (e.g. AACR2, CCO)
source	refers to the local, print, or electronic source from which information is derived for a specific element (e.g. Grove Dictionary of Art). Please note: SOURCE is also used as an element and should be used when you want to record a single print or electronic source for information pertaining to the entire record rather than pertaining to individual elements.
vocab	refers to the controlled vocabulary source from which the term or phrase is derived (e.g. AAT, LCSH).
xml:lang	refers to the language in which the information is recorded in the system (e.g. English, French).

Source: http://www.loc.gov/standards/vracore/VRA_Core4_Intro.pdf p.8

Example 1: VRA 4.0

VRA Core Element	XML element	*XML attribute*	Data example
WORK	Work		
		id	w_987654321
		refid	000347
		source	History of Art Visual Resources Collection, UCB
COLLECTION	Collection		
		id	c_876543210
		refid	BANC MSS 67/125 z
		source	Bancroft Library, UCB
IMAGE	Image		
		id	i_765432109
		refid	388438
		source	History of Art Visual Resources Collection, UCB

Example 2: VRA 4.0

VRA Core Element	**XML element**	**XML sub element**	***XML attribute***	**XML sub element**	**Data example (display value in bold)**
AGENT	**agent**				**School of Peter Paul Rubens (1577-1640)**
		Name			Rubens, Peter Paul
			type		Personal
			vocab		ULAN
			refid		500002921
		Culture			Flemish
		Dates			**1577 - 1640**
			type		life
				earliestDate	1577
				latestDate	1640
		Dates			**1577 - 1640**
			type		activity
				earliestDate	1590
				latestDate	1640
		Role			Painter (artist)
			vocab		AAT
			refid		300025136
		attribution			School of

Source: http://www.loc.gov/standards/vracore/VRA_Core4_Element_Description.pdf

Appendix J

7.1.13. Other resources from survey data

In response to the question "*Can you let us know what are your preferred online resources*", survey respondents ranked the listed resources on a scale of 5 to 1 and they input additional comments, as appropriate. Subsequent analysis of their comments showed that art experts consult diverse online resources (Table 7.1) when conducting their research, encompassing general art specialist resources such as academic literature, Google Images and auction house websites and specific online museums, depending on the particular artwork under research.

Table 7.1 Other online resources consulted by art specialists

Other' resources identified by domain experts
Art Forum
artprice
Art UK
ArtNet
Blouin ART Info
Bodleian
British Library
British Museum
commercial gallery sites
E-flux
Fine-arts-museum.be
Frans Hals Museum
Frieze
Gallica
Google books
Google images
Hyperallergic
Jstor
KIK-IRPA (balat)
Mauritshuis
MET
Metropolitan Museum of Art website
Morgan
museum websites
National Gallery London
National Portrait Gallery
NY and NGA, Washington, DC online entries.
Rijksmuseum
Sotheby's
TATE website
The Art Newspaper
VADS NICE Paintings.
VAI
Victoria & Albert

Appendix K

7.1.14. Museum systems online

Table 7.2 Artwork display: in online museum systems

Information displayed with an artwork	NG Ireland	IMMA	Hugh Lane	NG America	NG UK	MOMA	Louvre	Van Gogh	Met	RKD	**** Research Prototype
Image	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Artist											
Name(s)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Dates (Birth or Death)	Y	Y	Y	Y	Y	L	Y	Y	Y	L	Y
Artist Nationality	Y	Y	L	Y	L	L		L	Y	L	Y
About	L	Y	L	L	L	L	Y	L	L	L	L
Artwork											
Title(s)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Date Created	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Material/Technique	Y	Y	Y	Y	Y	Y		Y	Y	Y	Y
Measurements	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Inscription	Y				Y		Y		Y	Y	Y
Description/Entry	Y		Y	Y	Y	Y	Y	Y	Y	Y	Y
Entry or label	Y	Y		Y					Y		Y
Credit Line	Y	Y	Y		Y	Y		Y	Y		Y
Object number	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Catalogue Raisonné No.											Y
Exhibition History	Y			Y					Y	Y	Y
Provenance	Y			Y		Y			Y	Y	Y
Literature/Bibliography				Y			Y		Y	Y	Y
Acquisition		Y			Y	Y	Y				Y
Current Location				Y	Y	Y	Y	Y	Y	Y	Y
Place created/depicted					L	Y		Y			
On Display (Y/N)		Y		Y		Y	Y		Y		
Collection Name				Y		Y					
Copyright					L	Y		Y	Y		
Technical Details				Y	L				L		
Shape										Y	
References									Y		
Terms/keywords				24 Iconclass				Y		Y	
Related Content											
Audio				L		L	Y		L		
Artist				L							
images				L			Parts				
X-Ray images				L							
Infrared images					L		L		Y	Y	
Other materials								Study For		Y	
Related paintings/works				Y	L			Y	Y		
Exhibited with other paintings						L	L				
Artwork installation photos						Y					
Historical Text							Y		L		
Archives **Linked / Partially linked**							Y	L	List	Y	Y

B = Birth Year, L = available via a hyperlink on the page, VG Arch = Van Gogh Archives

Websites last visited 21/01/2019 – 24/01/2019

Table 7.3 Artwork search: in online museum systems

(Table continues on the next page)

Search Options	NG Ireland	IMMA	Hugh Lane	NG America	NG UK	MOMA	Louvre	Van Gogh	Met	RKD	**** Research Prototype
Default search box / free text	Y	Y		Y	Y	Y	Y	Y	Y	Y	
Artists A to Z		Y	Yo		Yo						
Title /Keyword in Artwork Title	Y	D	Yo	Y			Y		Yo	Y	Y
Collection Type (Subject area)	Y	D					Y			Y	Y
Date / Year	Y	D					Y			Y	Y
Style/Movement											Y
Technique										Y	Y
(Supporting) Medium	Y									Y	Y
Keywords object desc/info			Yo	Y					Yo	Y	Y
Object/Accession Number	Y		Yo	Y					Yo	Y	Y
Catalogue Raisonne				Y							Y
Artist name	Y	D	Yo	Y			Y		Yo	Y	Y
Artist Birth Year range											Y
Artist Nationality	Y										Y
Artist Gender											Y
Current Location							Y			Y	
Period (Century)	Y				Yo		Y			Y	
Type: Paintings, Artist or Events)					Yo						
Latest arrivals (Acquisitions)					Yo						
Latest arrivals (Loans)					Yo						
Credit Line				Y							
Provenance Name				Y							
Acquisition Method	Y										
Acquisition Date	Y										
Exhibition History				Y							
Online Edition Name				Y							
Gallery									Yo		
Reproduction number										Y	
Artistically related to other work										Y	
Artist related work										Y	
Location Made										Y	
Iconclass										Y	
Shape										Y	
Signature sort										Y	
Physical connections										Y	
Collection Name										Y	
Project										Y	
Technical Examination Type										Y	
Researcher										Y	
Document Type										Y	
Research Start Date										Y	

Y denotes that the option is provided.

Yo denotes that the option cannot be combined with another option when searching for an artwork.

D denotes that the user may select a value from a dropdown list

Websites last visited 08/02/2019

Table 7.3 Artwork search: in online museum systems

(Continued from previous page)

Search Options	NG Ireland	IMMA	Hugh Lane	NG America	NG UK	MOMA	Louvre	Van Gogh	Met	RKD	**** Research Prototype
Filtering Options											
Artist	Y							Y			
Artist Nationality				Y							
Collection Type / Category	Y			Y		Y		Y	Y		
Date / Year / Era	Y			Y		Y		Y	Y		
Medium	Y										
Acquisition Method	Y										
Provenance (Check Box)	Y										
Exhibition History (Check Box)	Y										
Online Editions				Y							
Styles				Y							
Location (where created)				Y				Y	Y		
Subjects (terms)				Y							
Alternate numbers				Y							
Genre								Y			
Department									Y		
Include or Exclude works											
Works on view ?		Y		Y		Y			Y		
Works with Images ?	Y	Y				Y			Y		
Works Out on loan ?		Y									
Works Online ?		Y		Y		Y					
Recent Acquisitions ?						Y					
Include Uncatloged Works ?						Y					
Highlights ?									Y		
Open access ?									Y		

Bibliography

Abiteboul, S., Manolescu, I., Rigaux, P., Rousset, M.-C., & Senellart, P. (2011). Ontologies, RDF, and OWL in *Web Data Management* (1 ed., pp. 143-170): Cambridge University Press.

Antoniou, G., Groth, P., Harmelen, F. V., & Hoekstra, R. (2012). *A Semantic Web Primer* (3rd ed.): MIT Press.

Antoniou, G., & Van Harmelen, F. (2004). Web ontology language: Owl *Handbook on ontologies* (pp. 67-92): Springer.

Audi, R. (1999). *The Cambridge Dictionary of Philosophy* (R. Audi Ed. 2 ed.). Cambridge, UK: Cambridge University Press.

Baca, M. (2004). Fear of Authority? Authority Control and Thesaurus Building for Art and Material Culture Information. *Cataloging & Classification Quarterly, 38*(3-4), 143-151. doi: 10.1300/J104v38n03_13

Baca, M. (2008a). *Introduction to Metadata* (M. Baca Ed. 2nd edition (15 Sep 2008) ed.): Getty Research Institute,U.S.; .

Baca, M. (2008b). News. *Visual Resources, 24*(2), 103-104. doi: 10.1080/01973760802042572

Baca, M. (2013). Multilingual AAT and Digital Art History Work at the Academia Sinica. *Visual Resources, 29*(4), 273-275. doi: 10.1080/01973762.2013.846776

Baca, M., & Harpring, P. (2014 -Revised online 2016). *Categories for the Description of Works of Art*. At: http://www.getty.edu/research/publications/electronic_publications/cdwa/definitions.pdf: J. Paul Getty Trust.

Baca, M., Harpring, P., Lanzi, E., McRae, L., & Whiteside, A. B. (2006). *Cataloging cultural objects: A guide to describing cultural works and their images*: American Library Association.

Bachman, C. W. (1969). Data structure diagrams. *SIGMIS Database, 1*(2), 4-10. doi: 10.1145/1017466.1017467

Bagley, P. R. (1968). Extension of Programming Language Concepts. *Available at http://www.dtic.mil/dtic/tr/fulltext/u2/680815.pdf, University City Science Centre, Philadelphia, Pennsylvania*, 223.

Bazeley, P. (2013). *Qualitative Data Analysis – Practical Strategies*. LA, London, New Delhi: Sage Publications Lit.

Beaudoin, J. E. (2005). Image and Text: A Review of the Literature Concerning the Information Needs and Research Behaviors of Art Historians*,* 34.

Beaudoin, J. E. (2007). Visual Materials and Online Access: Issues Concerning Content Representation*,* 24.

Beaudoin, J. E. (2014). A framework of image use among archaeologists, architects, art historians and artists. *Journal of Documentation, 70*(1), 119-147.

Becker, J. (1977). Review of Iconclass-An Iconographic Classification System by H. van de Waal; L. D. Couprie; R. H. Fuchs; E. Tholen. *Simiolus: Netherlands Quarterly for the History of Art, 9*(1), 45-47. doi: 10.2307/3780424

Bell, C. (1913). *Art* (5 ed.). New York: Frederick A Stokes Company.

Berger, J. (2008). *Ways of Seeing*. London: British Broadcasting Company and Penguin Books.

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific american, 284*(5), 28-37.

Bernstein, A., Hendler, J., & Noy, N. (2016). A New Look at the Semantic Web. *Communications of the ACM, 59*(9), 35-37. doi: 10.1145/2890489

Blackburn, S. (2014). *The Oxford Dictionary of Philosophy (2 rev. ed.)* (S. Blackburn Ed.): Oxford University Press.

Blei, D. M. (2012). Probabilistic Topic Models. *Communications of the ACM, 55*(4), 77-84. doi: 10.1145/2133806.2133826

Boem, F., Camara, G., Chuk, E., Quattri, F., Ratti, E., Sanfilippo, E. M., & Sojic, A. (2013). Diverse perspectives on ontology: A joint report on the First IAOA Interdisciplinary Summer School on Ontological Analysis. *Applied Ontology, 8*(1), 59-71.

Boles, F. (1982). Disrespecting original order. *The American Archivist, 45*(1), 26-32.

Brachman, R. J., & Levesque, H. J. (2004). *Knowledge Representation and Reasoning*: Morgan Kaufmann Publishers.

Breakell, S. (2015). Archival practices and the practice of archives in the visual arts. *Archives & Records, 36*(1), 1-5. doi: 10.1080/23257962.2015.1018151

Brilliant, R. (1988). How an art historian connects art objects and information. *Library Trends, 37*(2), 120-129.

Brinkman, B., Flick, C., Gotterbarn, D., Miller, K., Vazansky, K., & Wolf, M. J. (2017). Listening to Professional Voices: Draft 2 of the ACM Code of Ethics and Professional Conduct. *Communications of the ACM, 60*(5), 105.

Bryman, A. (2012). *Social research methods* (4th ed.): Oxford university press.

Buchanan, F., Capanni, N., & González-Vélez, H. (2015). Distributed aggregation of heterogeneous Web-based Fine Art Information: enabling multi-source accessibility and curation. *The Knowledge Engineering Review, 30*(2), 220-236. doi: http://dx.doi.org/10.1017/S0269888914000319

Bunnin, N., & Yu, J. (2004). *The Blackwell Dictionary of Western Philosophy*: Blackwell Publishing.

Calaresu, M., & Shiri, A. (2015). Understanding Semantic Web: a conceptual model. *Library Review, 64*(1/2), 82-100. doi: doi:10.1108/LR-09-2014-0097

Campbell, W. S., Campbell, J. R., West, W. W., McClay, J. C., & Hinrichs, S. H. (2014). *Semantic analysis of SNOMED CT for a post-coordinated database of histopathology findings* (Vol. 21).

Chen, P. (1976). The entity-relationship model-toward a unified view of data. *ACM Trans. Database Syst., 1*(1), 9-36. doi: 10.1145/320434.320440

Chortaras, A., Drosopoulos, N., Kollia, I., & Simou, N. (2014). Web Reasoning for Cultural Heritage. In R. Kontchakov & M.-L. Mugnier (Eds.), *Web Reasoning and Rule Systems* (Vol. 8741, pp. 23-28): Springer International Publishing.

Coburn, E., Lanzi, E., O'Keefe, E., Stein, R., & Whiteside, A. (2010). The Cataloging Cultural Objects experience: Codifying practice for the cultural heritage community. *IFLA Journal, 36*(1), 16-29. doi: 10.1177/0340035209359561

Codd, E. F. (1970). A relational model of data for large shared data banks. *Commun. ACM, 13*(6), 377-387. doi: 10.1145/362384.362685

Codd, E. F. (1971). *A data base sublanguage founded on the relational calculus*. Paper presented at the Proceedings of the 1971 ACM SIGFIDET (now SIGMOD) Workshop on Data Description, Access and Control, San Diego, California.

Codd, E. F. (1979). *Extending the data base relational model to capture more meaning*. Paper presented at the Proceedings of the 1979 ACM SIGMOD international conference on Management of data, Boston, Massachusetts.

Cohen, S. M., & Matthews, G. B. (1991). *Ammonius On Aristotle Categories*. London: Gerald Duckworth & Co. Ltd.

Concordia, C., Gradmann, S., & Siebinga, S. (2010). Not just another portal, not just another digital library: A portrait of Europeana as an application program interface. *IFLA Journal, 36*(1), 61-69. doi: 10.1177/0340035209360764

Cook, T. (1997). What is past is prologue: a history of archival ideas since 1898, and the future paradigm shift. *Archivaria, 42*.

Coronel, C., & Morris, S. (2016). *Database systems: design, implementation, & management* (12th ed.). USA: Cengage Learning.

Corrado, E. M. (2018). Discovery Products and the Open Archives Initiative Protocol for Metadata Harvesting. *International Information & Library Review, 50*(1), 47-53. doi: 10.1080/10572317.2017.1422905

Couprie, L. D. (1978). Iconclass, a device for the iconagraphical analysis of art objects. *Museum XXX*(3/4), 194-195.

Craig, E. (1998). Ontology. In E. Craig (Ed.), *Routledge Encyclopedia of Philosophy*. London: Routledge. Retrieved January 05, 2015, from http://www.rep.routledge.com/article/N039.

Creswell, J. W. (2014). *Research Design, Qualitative, Quantitative, & Mixed Methods Approaches* (4th ed.). California 91320: SAGE Publications, Inc.

Crofts, N., & Reed, P. (1996). CIDOC data modeling working group. *Archives and Museum Informatics, 10*(1), 89-91. doi: 10.1007/BF02802421

Crookham, A. (2015). Curatorial constructs: archives in fine art exhibitions (Vol. 36, pp. 18-28): Routledge.

Culleton, C. A. (2014). Competing Concepts of Culture: Irish Art at the 1924 Paris Olympic Games. *Estudios Irlandeses*(9), 24-34.

Daconta, M. C., Obrst, L. J., & Smith, K. T. (2003). *The Semantic Web*. Indianapolis, Indiana & Canada: Wiley Publishing Inc.

Danto, A. C. (2001). *The Transfiguartion of the Commonplace* (11 ed.). Cambridge, MA, USA & London, UK: Harvard University Press.

Davies, D. (2009). The Primacy of Practice in the Ontology of Art. *Review article, 67*(2), 159-171. doi: 10.1111/j.1540-6245.2009.01345.x

Davis, R., Shrobe, H., & Szolovits, P. (1993). What is a Knowledge Representation? . *AI Magazine, 14*(1), 17-33.

de Rynck, P. (2004). *How to Read a Painting: Decoding, Understanding and Enjoying the Old Masters*. London: Thames & Hudson.

Der Manuelian, P. (2013, Oct. 28 2013-Nov. 1 2013). *Giza 3D: Digital archaeology and scholarly access to the Giza Pyramids: The Giza Project at Harvard University.* Paper presented at the Digital Heritage International Congress (DigitalHeritage), 2013.

Devarakonda, R., Palanisamy, G., Green, J. M., & Wilson, B. E. (2011). Data sharing and retrieval using OAI-PMH. *Earth Science Informatics, 4*(1), 1-5. doi: 10.1007/s12145-010-0073-0

Devereux, D. (2003). *Plato: Metaphysics in The Blackwell Guide to Ancient Philosophy* (C. Shields Ed.). Malden USA, Oxford UK: Blackwell Publishing.

Dijkshoorn, C., Aroyo, L., Schreiber, G., Wielemaker, J., & Jongma, L. (2014). Using Linked Data to Diversify Search Results a Case Study in Cultural Heritage. In K. Janowicz, S. Schlobach, P. Lambrix & E. Hyvönen (Eds.), *Knowledge Engineering and Knowledge Management* (Vol. 8876, pp. 109-120): Springer International Publishing.

Dodd, J. (2012). Defending the Discovery Model in the Ontology of Art: A Reply to Amie Thomasson on the Qua Problem. *The British Journal of Aesthetics, 52*(1), 75-95. doi: 10.1093/aesthj/ayr047

Doerr, M. (2003). The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine, 24*(3), 75-92.

Doerr, M., Gradmann, S., Hennicke, S., Isaac, A., Meghini, C., & van de Sompel, H. (2010). *The europeana data model (edm).* Paper presented at the World Library and Information Congress: 76th IFLA general conference and assembly.

Dondis, D. A. (1973). *A Primer of Visual Literacy*: MIT Press.

Duchein, M. (1992). The History of European Archives and the Development of the Archival Profession in Europe. *The American Archivist, 55*(1), 14-25.

Edwards, P. N., Mayernik, M. S., Batcheller, A. L., Bowker, G. C., & Borgman, C. L. (2011). Science friction: Data, metadata, and collaboration. *Social Studies of Science, 41*(5), 667-690. doi: 10.1177/0306312711413314

Elmasri, R., & Navathe, S. B. (2014). *Fundamentals of Database Systems* (6th ed.): Pearson.

Elmasri, R., Weeldreyer, J., & Hevner, A. (1985). The category concept: An extension to the entity-relationship model. *Data & Knowledge Engineering, 1*(1), 75-116. doi: https://doi.org/10.1016/0169-023X(85)90027-8

Eslami, S., & Nazami, E. (2011, 11-13 March 2011). *An automatic approach for Topic Maps development using relational databases.* Paper presented at the 2011 3rd International Conference on Computer Research and Development.

Europeana Project. (2014). *Europeana Data Model – Mapping Guidelines v2.2* (pp. 46). Retrieved from http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Mapping_Guidelines_v2.2.pdf

Fagin, R. (1977). Multivalued dependencies and a new normal form for relational databases. *ACM Trans. Database Syst., 2*(3), 262-278. doi: 10.1145/320557.320571

Fagin, R. (1981). A normal form for relational databases that is based on domains and keys. *ACM Trans. Database Syst., 6*(3), 387-415. doi: 10.1145/319587.319592

Feigenbaum, G., & Reist, I. J. (2012). *Provenance: An Alternate History of Art* (G. Feigenbaum & I. J. Reist Eds.). Los Angeles: Getty Research Institute, Getty Publications.

Fensel, D., van Harmelen, F., Horrocks, I., McGuinness, D. L., & Patel-Schneider, P. F. (2001). OIL: an ontology infrastructure for the Semantic Web. *Intelligent Systems, IEEE, 16*(2), 38-45. doi: 10.1109/5254.920598

Fernández-López, M., Gómez-Pérez, A., & Suárez-Figueroa, M. C. (2013). Methodological guidelines for reusing general ontologies. *Data & Knowledge Engineering, 86*(0), 242-275. doi: http://dx.doi.org/10.1016/j.datak.2013.03.006

Fink, E. E. (1999). The Getty Information Institute A Retrospective. *D-Lib Magazine, 5*(3), 15. doi: 10.1045/march99-fink
Fonseca, F., & Martin, J. (2007). Learning The Differences Between Ontologies and Conceptual Schemas Through Ontology-Driven Information Systems. *Journal of the Association for Information Systems, 8*(2), 129-142.
Fowler Jr, F. J. (2014). *Survey research methods* (5 ed.): Sage publications.
Francisco-Revilla, L., Trace, C. B., Li, H., & Buchanan, S. A. (2014). Encoded archival description: Data quality and analysis. *Proceedings of the Association for Information Science and Technology, 51*(1), 1-10.
Gaehtgens, T. W. (2013). Thoughts on the Digital Future of the Humanities and Art History. *Visual Resources, 29*(1-2), 22-25. doi: 10.1080/01973762.2013.761110
Garbacz, P., & Trypuz, R. (2013). A metaontology for applied ontology. *Applied Ontology, 8*(1), 1-30.
Garrido, A. L., Buey, M. G., Escudero, S., Ilarri, S., Mena, E., & Silveira, S. B. (2013, 4-6 Nov. 2013). *TM-Gen: A Topic Map Generator from Text Documents.* Paper presented at the 2013 IEEE 25th International Conference on Tools with Artificial Intelligence.
Garshol, L. M. (2004). Metadata? Thesauri? Taxonomies? Topic maps! Making sense of it all. *Journal of Information Science, 30*(4), 378-391.
Georgiadis, H., Banos, V., Stathopoulou, I. O., Stathopoulos, P., Houssos, N., & Sachini, E. (2014, 7-9 July 2014). *Ensuring the quality and interoperability of open cultural digital content: System architecture and scalability.* Paper presented at the Information, Intelligence, Systems and Applications, IISA 2014, The 5th International Conference on.
Gombrich, E. H. (1995). *The Story of Art* (15th ed.). London: Phaidon Press Limited.
Gomez-Perez, A., & Corcho, O. (2002). Ontology languages for the Semantic Web. *IEEE Intelligent Systems, 17*(1), 54-60. doi: 10.1109/5254.988453
Grabar, N., Hamon, T., & Bodenreider, O. (2012). Ontologies and terminologies: Continuum or dichotomy? *Applied Ontology, 7*(4), 375-386.
Grimm, S., Pascal, H., & Abecker, A. (2007). Knowledge Representation and Ontologies in Semantic Web Services (pp. p51-105, 155p).
Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition, 5*(2), 199-220. doi: http://dx.doi.org/10.1006/knac.1993.1008
Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies, 43*(5–6), 907-928.
Guarino, N. (1995). Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human-Computer Studies, 43*(5–6), 625-640. doi: http://dx.doi.org/10.1006/ijhc.1995.1066
Guarino, N. (1998). Formal Ontology and Information Systems. Formal Ontology in Information

Systems, FOIS'98, Trento, Italy, IOS Press.
Guarino, N., Oberle, D., & Staab, S. (2009). What Is an Ontology? In S. Staab & R. Studer (Eds.), *Handbook on Ontologies* (pp. 1-17): Springer Berlin Heidelberg.
Guarino, N., & Welty, C. (2002). Evaluating Ontological Decisions with ONTOCLEAN. *Communications of the ACM, 45*(2), 61-65.
Harpring, P. (2010a). Development of the Getty Vocabularies: AAT, TGN, ULAN, and CONA. *Art Documentation: Bulletin of the Art Libraries Society of North America, 29*(1), 67-72.
Harpring, P. (2010b). *Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Works*. www.gettypublications.org 2010 by the Getty Research Institute, Los Angeles CA 90049-1682: Getty Publications, Gregory M. Britton.
Hedden, H. (2010). Taxonomies and controlled vocabularies best practices for metadata. *Journal Of Digital Asset Management, 6*(5), 279-284. doi: 10.1057/dam.2010.29
Heffernan, J. (2015). DIGITAL PUBLIC LIBRARY OF AMERICA AND EUROPEANA—two major history portals. *History Ireland*(1), 8.
Henriksson, J., Pradel, M., Zschaler, S., & Pan, J. (2008). Ontology Design and Reuse with Conceptual Roles. In D. Calvanese & G. Lausen (Eds.), *Web Reasoning and Rule Systems* (Vol. 5341, pp. 104-118): Springer Berlin Heidelberg.
Hepp, M. (2008). Ontologies: State of the Art, Business Potential, and Grand Challenges. In M. Hepp, P. Leenheer, A. Moor & Y. Sure (Eds.), *Ontology Management* (Vol. 7, pp. 3-22): Springer US.

Hildebrand, M., van Ossenbruggen, J., Hardman, L., & Jacobs, G. (2009). Supporting subject matter annotation using heterogeneous thesauri: A user study in Web data reuse. *International Journal of Human-Computer Studies, 67*(10), 887-902. doi: http://dx.doi.org/10.1016/j.ijhcs.2009.07.008

Hodge, S. (2014). *How to Look at Art*. London: Tate Publishing.

Hoehndorf, R., Dumontier, M., & Gkoutos, G. V. (2013). Evaluation of research in biomedical ontologies. *Briefings in Bioinformatics, 14*(6), 696-712.

Hofstadter, A., & Kuhns, R. (2009). *Philosophies of art and beauty: Selected readings in aesthetics from Plato to Heidegger*. USA: University of Chicago Press.

Hookway, C. (2010). Ontological Commitment. In J. Dancy, E. Sosa & M. Steup (Eds.), *A Companion to Epistemology* (2 ed., pp. 563-564). Malden, MA, USA, Oxford, UK: Wiley-Blackwell.

Horrocks, I. (2008). Ontologies and the semantic web. *Commun. ACM, 51*(12), 58-67. doi: 10.1145/1409360.1409377

Horrocks, I., Patel-Schneider, P. F., & van Harmelen, F. (2003). From SHIQ and RDF to OWL: the making of a Web Ontology Language. *Web Semantics: Science, Services and Agents on the World Wide Web, 1*, 7-26. doi: 10.1016/j.websem.2003.07.001

Hourihane, C. P. (2014). Classifying Subject Matter in Medieval Art: The Index of Christian Art at Princeton University. *Visual Resources, 30*(3), 255-262. doi: 10.1080/01973762.2014.936103

Hseih, H.-F., & Shannon, E. S. (2005). Three Approaches to Qualitative Content Analysis. *Qualitative Health Research, 15*(9), 1277-1288.

Hu, J., Lv, Y., & Zhang, M. (2014). The Ontology Design of Intangible Cultural Heritage Based on CIDOC CRM. *International Journal of U- & E-Service, Science & Technology, 7*(1), 261-274. doi: 10.14257/ijunesst.2014.7.1.24

Huhns, M. N., & Singh, M. P. (1997). Ontologies for agents. *Internet Computing, IEEE, 1*(6), 81-83. doi: 10.1109/4236.643942

Hyvönen, E. (2012). Publishing and Using Cultural Heritage Linked Data on the Semantic Web. *Synthesis Lectures on the Semantic Web: Theory and Technology, 2*(1), 1-159. doi: 10.2200/S00452ED1V01Y201210WBE003

Hyvönen, E., Mäkelä, E., Salminen, M., Valo, A., Viljanen, K., Saarela, S., . . . Kettula, S. (2005). MuseumFinland—Finnish museums on the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web, 3*(2–3), 224-241. doi: http://dx.doi.org/10.1016/j.websem.2005.05.008

ICA. (2000). ISAD(G): General International Standard Archival Description.

ICA. (2004). International Standard Archival Authority Record for Corporate Bodies, Persons and Families. *International Counctil of Archives, http://www.ica.org/en/public-resources/standards, Last accessed 14 Nov 2016*.

Irwin, T. (1995). The Theory of Forms *Plato's Ethics* (pp. 148-168). New York: Oxford University Press.

Isemann, D., & Ahmad, K. (2014). Ontological access to images of fine art. *J. Comput. Cult. Herit., 7*(1), 1-25. doi: 10.1145/2538030

Ivanović, M., & Budimac, Z. (2014). An overview of ontologies and data resources in medical domains. *Expert Systems with Applications, 41*(11), 5158-5166. doi: http://dx.doi.org/10.1016/j.eswa.2014.02.045

Jacquette, D. (2002). *Ontology*. Bucks HP5 1HQ: Acumen Publishing Limited.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational researcher, 33*(7), 14-26.

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of mixed methods research, 1*(2), 112-133.

Kennedy, R. (2007). Masquerade & Spectacle:The Circus and the Travelling Fair in the work of Jack B. Yeats. In N. G. o. Ireland (Ed.). Ireland: Published in conjunction with the Exhibition at NGI.

Kent, W. (1983). A simple guide to five normal forms in relational database theory. *Commun. ACM, 26*(2), 120-125. doi: 10.1145/358024.358054

King, J. L. (2015). Humans in Computing: Growing Responsibilities for Researchers. *Communications of the ACM, 58*(3), 31-33. doi: 10.1145/2723675

Klein, G. O., & Smith, B. (2010). Concept Systems and Ontologies: Recommendations for Basic Terminology. *Transactions of the Japanese Society for Artificial Intelligence = Jinko Chino Gakkai ronbunshi, 25*(3), 433-441. doi: 10.1527/tjsai.25.433

Kless, D., Milton, S., & Kazmierczak, E. (2012). Relationships and relata in ontologies and thesauri: Differences and similarities. *Applied Ontology, 7*(4), 401-428.
Kless, D., Milton, S., Kazmierczak, E., & Lindenthal, J. (2014). Thesaurus and ontology structure: Formal and pragmatic differences and similarities. *Journal of the Association for Information Science and Technology*, n/a-n/a. doi: 10.1002/asi.23268
Koide, N., Kubo, T., Nishida, S., Shibata, T., & Ikeda, K. (2015). Art Expertise Reduces Influence of Visual Salience on Fixation in Viewing Abstract-Paintings. *PLoS One, 10*(2), 1-14. doi: 10.1371/journal.pone.0117696
Koot, R. (2015). RKD Explore: searching for art-historical information in the 21st Century. *Art Libraries Journal, 40*(3), 21-27.
Koot, R., & Koppenol, K. (2016). From Marginal to Mainstream: Art Ephemera as Research Material at the RKD. *Art Documentation: Journal of the Art Libraries Society of North America, 35*(1), 43-56. doi: doi:10.1086/685976
Kothari, C. R. (2004). *Research methodology: Methods and techniques*: New Age International.
Kroenke, D., & Auer, D. J. (2012). *Database processing, Fundamentals, Design and Implementation* (12 Ed.): Pearson.
Kurtz, D., Parker, G., Shotton, D., Klyne, G., Schroff, F., Zisserman, A., & Wilks, Y. (2009, 9-11 Dec. 2009). *CLAROS - Bringing Classical Art to a Global Public.* Paper presented at the e-Science, 2009. e-Science '09. Fifth IEEE International Conference on.
Lagoze, C., & Van de Sompel, H. (2001). *The open archives initiative: building a low-barrier interoperability framework*. Paper presented at the Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries, Roanoke, Virginia, USA.
Lagoze, C., & Van de Sompel, H. (2003). The making of the Open Archives Initiative protocol for metadata harvesting. *Library Hi Tech, 21*(2), 11.
Lagoze, C., Van de Sompel, H., Nelson, M., & Warner, S. (2015). The Open Archives Initiative Protocol for Metadata Harvesting. from http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm
Lamarque, P. (2012). *Work and Object: Explorations in the Metaphysics of Art*. Oxford: Oxford University Press.
Lassila, O., & McGuinness, D. (2001). The Role of Frame-Based Representation on the Semantic Web.
LeBoeuf, P., Doerr, M., Ore, C. E., & Stead, S. (2013). Definition of the CIDOC Conceptual Reference Model V 5.1.2 October 2013. *available at http://www.cidoc-crm.org/docs/cidoc_crm_version_5.1.2.pdf*, 152.
Leder, H., Gerger, G., Brieber, D., & Schwarz, N. (2014). What makes an art expert? Emotion and evaluation in art appreciation. *Cognition and Emotion, 28*(6), 1137-1147. doi: 10.1080/02699931.2013.870132
Lessing, A. (1965). What Is Wrong with a Forgery? *The Journal of Aesthetics and Art Criticism, 23*(4), 461-471. doi: 10.2307/427668
Levine, J. A. (2009). The Importance of Provenance Documentation in the Market for Ancient Art and Artifacts: the Future of the Market May Depend on Documenting the Past. *Journal of Art, Technology & Intellectual Property Law, 19*(2), 219-233.
Levinson, J. Art, work of Grove Art Online. Oxford Art Online: Oxford University Press, accessed April 20, 2014. Retrieved from <http://www.oxfordartonline.com.elib.tcd.ie/subscriber/article/grove/art/T004287>.
Long, M., & Schonfeld, C. R. (2014). Preparing for the Future of Research Services for Art History. *Art Documentation: Journal of the Art Libraries Society of North America, 33*(2 (Fall 2014)), 192-205.
Lowe, E. J. (2001). *Opening plenary talk: Recent Advances in Metaphysics*. Paper presented at the Proceedings of the international conference on Formal Ontology in Information Systems - Volume 2001, Ogunquit, Maine, USA.
Lowe, E. J. (2005). Ontology. In T. Hondrich (Ed.), The Oxford Companion to Philosophy (2ed.): Oxford University Press. Retrieved from http://www.oxfordreference.com/10.1093/acref/9780199264797.001.0001/acref-9780199264797-e-1806. doi: 10.1093/acref/9780199264797.013.1806
Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014). *The Stanford CoreNLP natural language processing toolkit.* Paper presented at the Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations.

Margolis, J. (1977). The Ontological Peculiarity of Works of Art. *Journal of Aesthetics & Art Criticism, 36*(1), 45.
Margolis, J. (1979). Recent Currents in Aesthetics of Relevance to Contemporary Visual Artists. *Leonardo, 12*(2), 111-119. doi: 10.2307/1573833
Margolis, J., Currie, G., Martin, R. L., & Van Camp, J. Ontology of Art. *Encyclopedia of Aesthetics*.
Markman, A. B. (2013). *Knowledge representation*: Psychology Press.
Marty, P. F., Rayward, W. B., & Twidale, M. B. (2003). Museum informatics. *Annual Review of Information Science and Technology, 37*(1), 259-294.
McCarthy, J. (1980). Circumscription—A form of non-monotonic reasoning. *Artificial Intelligence, 13*(1–2), 27-39. doi: http://dx.doi.org/10.1016/0004-3702(80)90011-9
McCarthy, K. (2012). The Liffey Swim-Ireland's great Olympic artefact. *History Ireland, 20*(4), 26-27.
McGuinness, D. L., & Van Harmelen, F. (2004). OWL web ontology language overview. *W3C recommendation, 10*(10), 2004.
Mealy, G. H. (1967). Another Look at Data. *AFIPS Proceedings of the Fall Joint Computer Conference, 1967., 31*, 525-534. doi: http://doi.ieeecomputersociety.org/10.1109/AFIPS.1967.112

Moser, S. (2010). THE DEVIL IS IN THE DETAIL: Museum Displays and the Creation of Knowledge. *Museum Anthropology, 33*(1), 22-32. doi: 10.1111/j.1548-1379.2010.01072.x
Myers, M. D., & Venable, J. R. (2014). A set of ethical principles for design science research in information systems. *Information & Management, 51*(6), 801-809. doi: http://dx.doi.org/10.1016/j.im.2014.01.002
Nagel, L., & Miller, C. (2013). The Tesauro de Arte & Arquitectura and Tesauro Regional Patrimonial: Tools for Describing and Enhancing Access to Latin American Cultural Resources Online. *Getty Research Journal*(5), 149-156. doi: 10.2307/41825353
Neches, R., Fikes, R., Finin, T., Gruber, T., Patil, R., Senator, T., & Swartout, W. R. (1991). Enabling Technology for Knowledge Sharing. *AI Magazine, 12*(3), 21.
Needleman, M. (2002). The Open Archives Initiative. *Serials Review 00987913, 28*(2).
Noy, N. F., & McGuinness, D. L. (2001). Ontology Development 101: A Guide to Creating Your First Ontology. *Technical Report KSL-01-05, Stanford Knowledge Systems Laboratory, 2001.*, 25.
Nurseitov, N., Paulson, M., Reynolds, R., & Izurieta, C. (2009). Comparison of JSON and XML data interchange formats: a case study. *Caine, 9*, 157-162.
Otto, J. J. (2014). Administrative Metadata for Long-Term Preservation and Management of Resources. *Library Resources & Technical Services, 58*(1), 4-32.
Pal, J. K. (2010). Metadata initiatives and emerging technologies to improve resource discovery. *Annals of Library and Information Studies, 57*(1), 44-53.
Panofsky, E. (1972). *Studies in Iconology: Humanistic Themes in the Art of the Renaissance*. USA: Harper and Row.
Pappas, M., Rusinek, C. S., Baugh, G. A., Boykin, A. W., Cobine, C. J., Dunham, C. S., . . . Threatt, M. L. (2006). Best Free Reference Web Sites (Vol. 46, pp. 34-39): American Library Association.
Park, J. R., & Childress, E. (2009). Dublin Core metadata semantics: an analysis of the perspectives of information professionals. *Journal of Information Science, 35*(6), 727-739. doi: 10.1177/0165551509337871
Pepper, S. (2006). Topic Maps - Supporting both knowledge and information. *ISO Focus, 3*(3), 27-29.
Peroni, S., Tomasi, F., & Vitali, F. (2013). Reflecting on the Europeana Data Model. In M. Agosti, F. Esposito, S. Ferilli & N. Ferro (Eds.), *Digital Libraries and Archives* (Vol. 354, pp. 228-240): Springer Berlin Heidelberg.
Pinfield, S. (2002). Creating institutional e-print repositories. *Serials, 15*(3), 261.
Pitti, D. V. (1997). Encoded Archival Description: The Development of an Encoding Standard for Archival Finding Aids. *The American Archivist, 60*(3), 268-283.
Pitti, D. V. (1999). Encoded archival description: An introduction and overview. *New Review of Information Networking, 5*(1), 61-69. doi: 10.1080/13614579909516936

Prom, C. J., & Habing, T. G. (2002). *Using the open archives initiative protocols with EAD*. Paper presented at the Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries, Portland, Oregon, USA.
Prom, C. J., Rishel, C. A., Schwartz, S. W., & Fox, K. J. (2007). *A unified platform for archival description and access*. Paper presented at the Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries, Vancouver, BC, Canada.
Pyle, H. (1989). *Jack B. Yeats, A Biography* (2nd ed.). New York: Barnes and Noble Books.
Pyle, H. (1992). *Jack B. Yeats, A Catalogue Raisonné of the Oil Paintings* (Vol. 1). London: André Deutsch Limited.
Pyle, H. (2001). To be loved as a cupboard: the Yeats Museum in the National Gallery of Ireland. *Eire-Ireland, 36*(3/4), 212-225.
Quine, W. V. (1948). On What There Is. *The Review of Metaphysics, 2*(5), 21-38. doi: 10.2307/20123117
Rai, A. (2017). Avoiding Type III Errors: Formulating IS Research Problems that Matter. *MIS Quarterly, 41*(2), iii-vii.
Rector, A. L., Brandt, S., & Schneider, T. (2011). Getting the foot out of the pelvis: modeling problems affecting use of SNOMED CT hierarchies in practical applications. *Journal of the American Medical Informatics Association, 18*(4), 432-440.
Reed, M. (2017). From the Archive to Art History. *Art Journal, 76*(1), 121-128. doi: 10.1080/00043249.2017.1332900
Reed, V. (2013). Due Diligence, Provenance Research, and the Acquisition Process at the Museum Of Fine Arts, Boston. *Journal of Art, Technology & Intellectual Property Law, 23*(2), 363-373.
Riley, J. (2017). Understanding Metadata: What is Metadata, and What is it For?: A Primer. *National Information Standards Organization (NISO), 3600 Clipper Mill Road, Suite 302, Baltimore, MD 21211, Available at:* http://www.niso.org/publications/understanding-metadata-what-metadata-and-what-it-primer.
Rorty, A. (2014). Dialogues with Paintings: Notes on How to Look and See. *The Journal of Aesthetic Education*(1), 1.
Rose, T. (2002). Technology's impact on the information-seeking behavior of art historians. *Art Documentation*, 35-42.
Rosenthal, T. G. (2003). *The Art of Jack B. Yeats* (2nd ed.). London: Andre Deutsch.
Rusch-Feja, D. (2002). The Open Archives Initiative and the OAI Protocol for Metadata Harvesting: rapidly forming a new tier in the scholarly communication infrastructure. *learned Publishing, 15*(3), 179-186.
Sandis, C. (2016). An Honest Display of Fakery: Replicas and the Role of Museums. . *Royal Institute of Philosophy Supplement, 79*, 241-259. doi: 10.1017/S1358246116000163
Schwartz, J. M., & Cook, T. (2002). Archives, records, and power: The making of modern memory. *Archival Science, 2*(1-2), 1-19.
Sekaran, U., & Bougie, R. (2016). *Research methods for business: A skill building approach*: John Wiley & Sons.
Sen, A. (2004). Metadata management: past, present and future. *Decision Support Systems, 37*(1), 151-173. doi: http://dx.doi.org/10.1016/S0167-9236(02)00208-7
Severance, C. (2012). Discovering JavaScript Object Notation. *Computer, 45*(4), 6-8. doi: 10.1109/MC.2012.132
Shepherd, E., & Pringle, R. (2002). Mapping Descriptive Standards Across Domains: a comparison of ISAD(G) and SPECTRUM. *Journal of the Society of Archivists, 23*(1), 17-34. doi: 10.1080/00379810220120537
Silverman, D. (2011). *Interpreting qualitative data: A guide to the principles of qualitative research* (4 ed.): SAGE Publications Limited.
Silvermann, A. (2014). Plato's Middle Period Metaphysics and Epistemology. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Fall 2014 ed.).
Sjöholm, J. (2014). The art studio as archive: tracing the geography of artistic potentiality, progress and production. *Cultural Geographies, 21*(3), 505-514. doi: 10.1177/1474474012473060
Smith, B. (1997). ON SUBSTANCES, ACCIDENTS AND UNIVERSALS: IN DEFENCE OF A CONSTITUENT ONTOLOGY. *Philosophical Papers, 26*(1), 105-127. doi: 10.1080/05568649709506558

Smith, B. (2004a). *Beyond concepts: ontology as reality representation.* Paper presented at the Proceedings of the third international conference on formal ontology in information systems (FOIS 2004), Turin.
Smith, B. (2004b). Ontology. In L. Floridi (Ed.), *The Blackwell guide to the philosophy of computing and information* (pp. 155-166). Malden, MA, USA & Oxford UK & Carlton, Victoria, Australia: Blackwell Publishing Ltd.
Smith, J. M., & Smith, D. C. P. (1977). Database abstractions: aggregation. *Commun. ACM, 20*(6), 405-413. doi: 10.1145/359605.359620
Sommerville, I. (2011). *Software engineering* (9th ed.). USA: Addison-Wesley.
Sowa, J. F. (2000). *Knowledge Representation: Logical, Philosophical and Computational Foundations*. USA: Brooks/Cole.
Stahl, B. C., Eden, G., Jirotka, M., & Coeckelbergh, M. (2014). From computer ethics to responsible research and innovation in ICT: The transition of reference discourses informing ethics-related research in information systems. *Information & Management, 51*(6), 810-818. doi: http://dx.doi.org/10.1016/j.im.2014.01.001
Stam, D. C. (1997). How Art Historians Look for Information. *Art Documentation: Journal of the Art Libraries Society of North America, 16*(2), 27-30. doi: 10.2307/27948896
Stathopoulou, I. O., Stathopoulos, P., Georgiadis, H., Houssos, N., Banos, V., & Sachini, E. (2014, 8-12 Sept. 2014). *An Open Cultural Digital Content Infrastructure.* Paper presented at the Digital Libraries (JCDL), 2014 IEEE/ACM Joint Conference on.
Steiner, K., Essmayr, W., & Wagner, R. (2001). *Topic maps-an enabling technology for knowledge management.* Paper presented at the 12th International Workshop on Database and Expert Systems Applications.
Strohmaier, M., Walk, S., Pöschko, J., Lamprecht, D., Tudorache, T., Nyulas, C., . . . Noy, N. F. (2013). How ontologies are made: Studying the hidden social dynamics behind collaborative ontology engineering projects. *Web Semantics: Science, Services and Agents on the World Wide Web, 20*(0), 18-34. doi: http://dx.doi.org/10.1016/j.websem.2013.04.001
Studtmann, P. (2011). Aristotle's Early and Late Ontologies. *Philosophy Compass, 6*(7), 469.
Sun, S., Liu, D., & Li, G. (2012). The Application of a Hierarchical Tree Method to Ontology Knowledge Engineering. *International Journal of Software Engineering & Knowledge Engineering, 22*(4), 571-593. doi: 10.1142/S0218194012500155
Swartout, W., & Tate, A. (1999). Ontologies. *Intelligent Systems and their Applications, IEEE, 14*(1), 18-19. doi: 10.1109/MIS.1999.747901
Sweeney, S. (2008). The Ambiguous Origins of the Archival Principle of "Provenance". *Libraries & the Cultural Record, 43*(2), 193-213.
Szekely, P., Knoblock, C., Yang, F., Zhu, X., Fink, E., Allen, R., & Goodlander, G. (2013). Connecting the Smithsonian American Art Museum to the Linked Data Cloud. In P. Cimiano, O. Corcho, V. Presutti, L. Hollink & S. Rudolph (Eds.), *The Semantic Web: Semantics and Big Data* (Vol. 7882, pp. 593-607): Springer Berlin Heidelberg.
Szekely, P., Knoblock, C. A., Yang, F., Fink, E. E., Gupta, S., Allen, R., & Goodlander, G. (2014). Publishing the Data of the Smithsonian American Art Museum to the Linked Data Cloud. *Journal of Humanities & Arts Computing: A Journal of Digital Humanities, 8*, 152-166. doi: 10.3366/ijhac.2014.0104
Teorey, T. J., Yang, D., & Fry, J. P. (1986). A logical design methodology for relational databases using the extended entity-relationship model. *ACM Comput. Surv., 18*(2), 197-222. doi: 10.1145/7474.7475
Thomasson, A. L. (2004). The Ontology of Art. In P. Kivy (Ed.), *The Blackwell Guide to Aesthetics*. Malden, MA: Blackwell.
Thomasson, A. L. (2005). The Ontology of Art and Knowledge in Aesthetics. *Journal of Aesthetics & Art Criticism, 63*(3), 221-230. doi: 10.1111/j.0021-8529.2005.00202.x
Thomasson, A. L. (2010). Ontological Innovation in Art. *Journal of Aesthetics & Art Criticism, 68*(2), 119-130. doi: 10.1111/j.1540-6245.2010.01397.x
Thompson, J. (2006). *How to Read a Modern Painting: understanding and enjoying the modern masters* London: Thames & Hudson Ltd.
Trapp, J. B. (1976). Iconclass: An Iconographic Classification by H. van de Waal. *The Burlington Magazine, 118*(884), 778. doi: 10.2307/878597
Tröndle, M., Greenwood, S., Bitterli, K., & van den Berg, K. (2014). The effects of curatorial arrangements. *Museum Management and Curatorship, 29*(2), 140-173. doi: 10.1080/09647775.2014.888820

Tzortzi, K. (2014). Movement in museums: mediating between museum intent and visitor experience. *Museum Management and Curatorship, 29*(4), 327-348. doi: 10.1080/09647775.2014.939844

Uschold, M., & Gruninger, M. (1996). Ontologies: principles, methods and applications. *The Knowledge Engineering Review, 11*(02), 93-136. doi: doi:10.1017/S0269888900007797

Van Hooland, S., & Verborgh, R. (2014). *Linked Data for Libraries, Archives and Museums: How to clean, link and publish your metadata*: Facet publishing.

van Hooland, S., Verborgh, R., De Wilde, M., Hercher, J., Mannens, E., & Van de Walle, R. (2013). Evaluating the success of vocabulary reconciliation for cultural heritage collections. *Journal of the American Society for Information Science and Technology, 64*(3), 464-479. doi: 10.1002/asi.22763

VanStraten, R. (1986). Panofsky and ICONCLASS. *Artibus et Historiae, 7*(13), 165-181. doi: 10.2307/1483254

Vassallo, V., Athanasiou, E., Hermon, S., & Eliades, I. (2013, Oct. 28 2013-Nov. 1 2013). *Publishing cultural heritage content for digital libraries: The case of the collections of the Byzantine Museum and Art Gallery of the Archbishop Makarios III foundation.* Paper presented at the Digital Heritage International Congress (DigitalHeritage), 2013.

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the Qualitative–Quantitative Divide: Guidelines for Conducting Mixed Methods Research in Information Systems. *MIS Quarterly, 37*(1), 21-54.

Venkatesh, V., Brown, S. A., & Sullivan, Y. W. (2016). Guidelines for Conducting Mixed-methods Research: An Extension and Illustration. *Journal of the Association for Information Systems, 17*(7), 435-495.

VRA. (2007). Visual Resource Association Core 4.0 Element Description *Available at: https://www.loc.gov/standards/vracore/VRA_Core4_Element_Description.pdf*.

W3C, O. W. G. (2012). OWL 2 Web Ontology Language Document Overview (Second Edition). https://www.w3.org/TR/2012/REC-owl2-overview-20121211/

Wahyuni, D. (2012). The Research Design Maze: Understanding Paradigms, Cases, Methods and Methodologies. *Journal of Applied Management Accounting Research, 10*(1), 69-80.

Waiboer, A. E. (2012). *Gabriel Metsu Life and Work A Catalogue Raisonné*. New Haven and London: Yale University Press.

Waltinger, U., Mehler, A., & Heyer, G. (2008). *Towards Automatic Content Tagging-Enhanced Web Services in Digital Libraries using Lexical Chaining.* Paper presented at the WEBIST (2).

Walton, K. L. (1970). Categories of Art. *The Philosophical Review, 79*(3), 334-367. doi: 10.2307/2183933

Weh, M. (2010). Production Determines Category: An Ontology of Art. *The Journal of Aesthetic Education, 44*(1), 84-99.

Weibel, S. L., & Lagoze, C. (1997). An element set to support resource discovery: The state of the Dublin Core: January 1997. *International Journal on Digital Libraries, 1*(2), 176.

Weinberg, B. H. (1995). Art & Architecture Thesaurus/Guide to Indexing and Cataloging With the Art & Architecture Thesaurus. *Journal of the American Society for Information Science, 46*(2), 152-160.

Wibisono, A., Koning, R., Grosso, P., Belloum, A., Bubak, M., & de Laat, C. (2013). OIntEd: online ontology instance editor enabling a new approach to ontology development. *Software: Practice and Experience, 43*(11), 1319-1335. doi: 10.1002/spe.2141

Wilson, R., O'Sullivan, A., Klassnik, R., Mulholland, H., O'Reilly, P. M., Cuív, R. Ó., & Orange, M. (1995). Vox Pop: Curate's Egg. *Circa*(72), 38-40. doi: 10.2307/25562813

Winston, P. (1992). Artificial Intelligence Addison-Wesley Pub Co.

Wölfflin, H. (1932). *Principles of Art History. The Problem of the Development of Style in Later Art* (Translated from 7th German Edition (1929) into English by M D Hottinger ed.). New York Dover Publications, Incorporated.

Yin, R. K. (2014). *Case Study Research Design and Methods* (5th ed.). California 91320: Sage Publications, Inc.

Yuta, S., Masaomi, K., & Yuki, T. (2017, 20-22 Oct. 2017). *A Native Topic Map Database Management System and Its N-gram Index for Efficient Search of Topics.* Paper presented at the 2017 International Symposium on Computer Science and Intelligent Controls (ISCSIC).

Zapounidou, S., Sfakakis, M., & Papatheodorou, C. (2014). *Integrating library and cultural heritage data models: the BIBFRAME - EDM case*. Paper presented at the Proceedings of the 18th Panhellenic Conference on Informatics, Athens, Greece.

Zoller, G., & DeMarsh, K. (2013). For the Record: Museum Cataloging from a Library and Information Science Perspective. *Art Documentation: Journal of the Art Libraries Society of North America, 32*(1), 54-70. doi: 10.1086/669989