Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

School of Computer Science & Statistics

ADAPT Research Centre for AI-Driven Digital Content Technology

Towards a metrics framework to facilitate the evaluation of temporal data modelling approaches in knowledge graphs

Sepideh Hooshafza

Supervisor:

Professor Gaye Stephens

A thesis submitted to the

University of Dublin, Trinity College Dublin

in fulfilment of the requirements of the degree of

Doctor of Philosophy

2025

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:

Sepideh Hooshafza

Acknowledgements

This thesis would not have been possible without the guidance, support, and encouragement of many individuals, and I would like to take this opportunity to express my heartfelt gratitude.

First and foremost, I extend my deepest appreciation to my supervisor, Gaye Stephens, whose expertise, insightful feedback, and unwavering support have been invaluable throughout the research and writing process. Your patience, constructive criticism, and consistent belief in my abilities have guided me through the many challenges of this project.

I also wish to acknowledge the Health Information and Quality Authority (HIQA), the ADAPT Research Centre, FAIRVASC Research Group, and the School of Computer Science and Statistics at Trinity College Dublin for providing the resources and a supportive environment that made this research possible. A special thanks goes to Dr Louise McQuaid and Dr Laura Keyes from HIQA for their assistance and encouragement in the first three years of my PhD.

To my colleagues in the ADAPT Research Centre, especially Dr Beyza Yaman, thank you for your advice and encouragement.

To my parents, Hadi and Parvaneh, my sister, Azadeh, and my brother, Amin—my deepest gratitude for your endless love, patience, and understanding throughout this entire process. Your constant support has been my source of strength, and your belief in me has fuelled my determination.

Lastly, to my husband, Navid—your love, understanding, and steadfast support have been invaluable during every phase of this PhD journey. You were by my side through the most challenging moments, and without your encouragement, I would not have been able to overcome the obstacles along the way. Thank you for always being there.

Abstract

In recent years, Knowledge Graphs (KGs) have emerged as a powerful tool for modelling data, utilised by companies like Facebook, Amazon, and IBM in applications such as natural language processing, recommendation systems, and question answering machines. However, modelling temporal data in Resource Description Framework (RDF)-based KGs remains a challenge, as RDF lacks a standardised way to represent time. This temporal dimension is crucial for domains such as healthcare, where time-varying data is essential for effective decision-making.

This thesis addresses the challenge of temporal data modelling in KGs by proposing the TEDME-KG metrics framework, a tool designed to evaluate various temporal data modelling approaches in RDF-based KGs. The research is grounded in a real-world healthcare use case focusing on medication management for patients with ANCA vasculitis, a rare disease. The use case leverages the FAIRVASC project, which involves the standardisation of vasculitis patient data across European registries.

The TEDME-KG metrics framework was developed using a structured methodology, incorporating expert feedback and literature review. It was evaluated based on its appropriateness, usability, and applicability through both a survey of Semantic Web researchers and its application to real-world temporal data modelling approaches.

Key findings include the successful development of the TEDME-KG metrics framework, which comprises three goals, three metric categories, 7 metrics subcategories, and 16 metrics offering a systematic approach to assessing the modelling, uplifting, and querying of temporal data in KGs. The framework was validated through its application in the healthcare domain, where it evaluated different temporal data models, contributing to the enhancement of the FAIRVASC ontology. In addition to the metrics framework, the thesis also presents a quality assessment framework for healthcare data sources and a categorised list of temporal data modelling approaches, offering valuable insights for both researchers and practitioners.

The results demonstrate that the TEDME-KG metrics framework provides a practical tool for evaluating temporal data modelling approaches, advancing the field by offering a structured method to address the complexities of temporal data in KGs. The evaluation results indicated that the TEDME-KG metrics framework is highly appropriate for its intended purpose, providing relevant and comprehensive metrics for temporal data

modelling. In terms of general usability, Semantic Web researchers found the framework easy to understand and apply, with minimal complexity. Regarding applicability, the framework was demonstrated to be effective in real-world scenarios, particularly in healthcare, where it supported the evaluation of temporal data models in medication data.

The TEDME-KG metrics framework, has shown its potential to be applied across a wide range of domains beyond healthcare, offering researchers and practitioners a systematic method to assess and improve temporal data models in knowledge graphs.

Contents

Chapter 1 Introduction 1

1.1 Motivation and scope 1

1.1.1 Healthcare Use case 3

1.2 Research Question 4

1.3 Research Objectives 5

1.3.1 Research Objective 1 (RO1) 5

1.3.2 Research Objective 2 (RO2) 5

1.3.3 Research Objective 3 (RO3) 6

1.3.4 Research Objective 4 (RO4) 6

1.4 Research Methodology 6

1.4.1 Methodology to achieve RO1 6

1.4.2 Methodology to achieve RO2 7

1.4.3 Methodology to achieve RO3 and RO4 7

1.5 Thesis contribution 10

1.5.1 Major contribution 10

1.5.2 Minor contribution 1 11

1.5.3 Minor contribution 2 11

1.5.4 Minor contribution 3 11

1.6 Impact 11

1.7 Publications 12

1.7.1 Published papers. 12

1.7.2 Publication Plan 14

1.8 Thesis Overview 15

1.8.1 Chapter 2 Problem Context Identification 15

1.8.2 Chapter 3 State of the Art 15

1.8.3 Chapter 4 Design 15

1.8.4 Chapter 5 Evaluation 16

1.8.5 Chapter 6 Implementation 16

1.8.6 Chapter 7 Conclusion 16

Chapter 2 Problem Context Identification 17

2.1 Data source quality assessment framework 17

2.1.1 Background in data source quality assessment in healthcare 17

2.1.2 Purpose 19

2.1.3 Methods 20

2.1.4 Phase 1: Literature review 21

2.1.5 Phase 2: Expert advisory groups 25

2.1.6 Phase 3: Framework design 26

2.1.7 Results 26

2.1.8 Discussion and conclusion 30

2.2 Time and temporal data in data and data source quality assessment 34

2.3 FAIRVASC Use case 36

2.4 Chapter Summary 38

Chapter 3 State of the art review 40

3.1 Temporal data modelling in healthcare 40

3.1.1 Motivation and scope 40
3.1.2 Method 42
3.1.3 Results 45
3.1.4 Discussion 48
3.1.5 Identified limitations 49

3.2 Systematic review 50

3.2.1 Methods 50
3.2.2 Results 52
3.2.3 Discussion 63

3.3 Time Ontologies 65

3.3.1 OWL-Time Ontology 65
3.3.2 W3C Data Cube Vocabulary 65
3.3.3 W3C Semantic Sensor Network (SSN) Ontology 66

3.4 FAIR Data Principles 67

3.5 Ontology Design Patterns (ODPs) 67

3.6 R2RML (RDB to RDF Mapping Language) 68

3.7 Evaluation approaches in KGs 69

3.7.1 Benchmarking RDF and SPARQL Systems 70
3.7.2 Evaluation of RDF Modelling Approaches 71
3.7.3 Evaluation of Temporal RDF Models 72
3.7.4 Identified limitations 73

3.8 Chapter Summary 74

Chapter 4 Design 76

4.1 Methodology Overview 76

4.2 Phase 1: Literature Review 77

4.2.1 Literature review on metrics for evaluating data modelling approaches in KGs – (S1) 78
4.2.2 Survey papers on temporal data modelling approaches in KGs - (S2) 79
4.2.3 Systematic review on temporal data modelling in KGs – (S3) 79
4.2.4 Data extraction and synthesis process 79
4.2.5 Literature review results 80
4.2.6 Thematic analysis 80
4.2.7 Thematic Analysis results 80
4.2.8 Metrics framework- First iteration 82

4.3 Phase 2: Updating metrics framework through GQM approach. 83

4.3.1 Pilot Study 84
4.3.2 Framework Refinement 85
4.3.3 Experiment Results 87
4.3.4 Detailed description of the metrics 90

4.4 Visual representation of TEDME_KG Metrics Framework 96

4.5 Chapter Summary 97

Chapter 5 Evaluation 99

5.1 Strategy 99

5.2 Sample Size 99

5.3 Evaluation of appropriateness of the TEDME-KG metrics framework 99

5.3.1 Results and interpretation 100

5.3.2 Overall Analysis of appropriateness 102

5.4 Validation of TEDME-KG metrics framework design 102

5.4.1 Suggestions for New Goals and Questions 103

5.4.2 Suggestions for New Metrics, Metrics Categories, or Metrics Subcategories 103

5.4.3 Suggestions for Grouping and Labelling of Metrics 103

5.4.4 Results and interpretation 104

5.4.5 Overall Analysis of design validation 106

5.5 Evaluating general usability 106

5.5.1 SUS score 106

5.5.2 Procedure 109

5.5.3 Results 110

5.5.4 Overall Analysis of general usability 112

5.5.5 Limitations 112

5.6 Chapter Summary 113

Chapter 6 Implementation 114

6.1 Methodology 114

6.1.1 Phase 1: Data Modelling Approaches Identification 115

6.1.2 Phase 2: Use Case Identification 115

6.1.3 Phase 3: KG Creation 116

6.1.4 Phase 4: KG Hosting 116

6.1.5 Phase 5: KG deployment 117

6.1.6 Phase 6: Metrics Identification and Evaluation 117

6.2 Technical Details 117

6.2.1 Experiment setup 117

6.2.2 Ontologies 117

6.2.3 R2RML Mappings 124

6.2.4 SPARQL Queries 125

6.2.5 Discussion on SPARQL queries 133

6.3 Evaluation results 133

6.3.1 Modelling metrics category 134

6.3.2 Uplifting Metrics Category 137

6.3.3 Querying Metrics Category 140

6.4 Chapter Summary 144

6.4.1 Comparison with baselines methods 144

6.4.2 Experiment Limitations 146

Chapter 7 Conclusion 147

7.1 Achievement of Research Objectives 147

7.2 Contributions 151

7.3 Impact and Uptake 153

7.4 Future Work 154

7.5 Final Remarks 156

References 158

Appendixes 166

List of Figures

Figure 1- The Goal-Question-Metric approach – (Adapted from Basili et al) 10
Figure 2- Development phases of data source quality assessment framework in healthcare settings 21
Figure 3-Flow chart of literature screening for review of assessment criteria relating to the quality of data sources in healthcare 24
Figure 4- Themes identified in the literature review on data source quality assessment criteria in healthcare 26
Figure 5- Data source quality assessment framework in healthcare settings 30
Figure 6- Base model - ontology without including medication time 37
Figure 7-Flow chart of literature screening for review of temporal data modelling in healthcare. 45
Figure 8-Flow chart of literature screening for systematic review of temporal data modelling in KGs 53
Figure 9- Temporal data modelling approaches in KGs in terms of scope of application 60
Figure 10- Design and evaluation phases of TEDME-KG metrics framework 77
Figure 11- Adding medication time to the base model based on Standard Reification approach (Left side) and Named Graph approach (Right side) 84
Figure 12- Visual representation of TEDME-KG Metrics Framework 96
Figure 13- Evaluation of the appropriateness of the TEDME-KG metrics Framework based on a set of criteria. 101
Figure 14- SUS score and its associated grading scales 108
Figure 15- The phases for evaluating temporal data modelling approaches in KGs 115
Figure 16- Adding the medication start and stop date to the Named Graph approach 118
Figure 17- Adding the medication start and stop date to the Standard Reification approach. 119
Figure 18- Adding the medication start and stop date to the RDF-Star approach. 120
Figure 19- Adding the medication start and stop date to the Temporal RDF Graph approach 121
Figure 20- Adding the medication start and stop date to the RDFMT approach 122
Figure 21- Adding the medication start and stop date to the RDFt approach. 123
Figure 22- Adding the medication start and stop date to the Singleton Property approach 124
Figure 23- Modelling complexity based on different approaches. 137
Figure 24- Uplifting complexity per approach 139
Figure 25- RDF Data storage space in KB per approaches 140
Figure 26- Data load time per approaches. 141
Figure 27- Query length per temporal question categories. 142
Figure 28- Query response time per approaches per temporal question category 143
Figure 29- Ontology without including medication time. 203
Figure 30- Modelling medication time using “Singleton Property” approach. 203
Figure 31- Modelling medication time using “Standard Reification”. 203

List of tables

Table 1- Search concepts and keywords for use in a literature review to identify areas of quality relating to data sources in healthcare settings 22
Table 2- Data source quality assessment framework - Sub themes definitions 28
Table 3- Search terms for a systematic review on modelling temporal data in healthcare settings 43
Table 4-Approaches of temporal data modelling in healthcare in relation to applications in healthcare in terms of functionality 48
Table 5-Search terms for a systematic review on modelling temporal data in knowledge graphs 51
Table 6- Comparison between approaches of temporal data modelling in knowledge graphs 54
Table 7- Studies related to the evaluation of temporal data modelling approaches in KGs 72
Table 8- Search concepts and keywords for use in a literature review to identify metrics relating to the evaluation of RDF data modelling approaches 79
Table 9- Thematic statements for the evaluation phases 81
Table 10- Metrics framework for evaluating temporal data modelling approaches in KGs – First Iteration 83
Table 11- TEDME-KG Metrics framework -2nd iteration 86
Table 12- Evaluation Results of Named Graph and Standard Reification based on TEDME-KG Metrics Framework 89
Table 13- Metrics details used in TEDME-KG Metrics framework. 90
Table 14- Metrics Categories and subcategories definitions 92
Table 15- Distribution of participants in the survey based on years of experience. 100
Table 16- System Usability Scale Questionnaire 107
Table 17- Amended version of the SUS questionnaire items used in this evaluation. 109
Table 18- Descriptive statistics results 111
Table 19- Temporal question categories 116
Table 20- Acronyms for each temporal data modelling approaches 133
Table 21- Implementation support and standardisation 134
Table 22- The results of complexity per temporal data modelling approaches 135
Table 23- Implementation support and standardisation-Uplifting category 138
Table 24- Uplifting Complexity results 138
Table 25- RDF data storage space in KB 139
Table 26- Implementation support and standardisation per approaches in Querying metrics category 140
Table 27-Data load time per approaches 141
Table 28- Query response time per approaches per temporal question categories 143
Table 29-Sample data for a patient using medication in a time interval 202
Table 30- Results of applying metrics framework to approach 1 and 2 204

Listings

Listing 1- Snippet of the R2RML code based on the Named Graph approach 124
Listing 2- Sparql query for the Named Graph approach for the explicit temporal (time point) category 125
Listing 3- Sparql query for the RDF—Star approach for the explicit temporal (time point) category 126
Listing 4- Sparql query for the Named Graph approach for the explicit temporal (time interval) category 127
Listing 5- Sparql query for the Standard Reification approach for the Implicit temporal (time point) category 127
Listing 6- Sparql query for the Standard Reification approach for the Implicit temporal (time interval) category 128
Listing 7- Sparql query for the RDF-Star approach for the Implicit temporal (time interval) category 128
Listing 8- Sparql query for the Singleton Property approach for the ordinal temporal (time point) category 129
Listing 9- Sparql query for the RDF-Star approach for the ordinal temporal (time point) category 129
Listing 10- Sparql query for the Singleton Property approach for the ordinal temporal (time interval) category 131
Listing 11- Sparql query for the Temporal RDF Graph approach for the temporal answer (time point) category 131
Listing 12- Sparql query for the Temporal RDF Graph approach for the temporal answer (time interval) category 132
Listing 13- Sparql query for the RDF-Star approach for temporal answer (time interval) category 132
Listing 14- Sparql query for calculating metrics related to the complexity of modelling. 135

Chapter 1 Introduction

1.1 Motivation and scope

In recent years, Knowledge Graphs (KGs) have drawn attention from academia and industry as a method for modelling data [1, 2]. KGs are defined as entities (nodes) and relationships (edges) in a semantic network [3]. They are used in a variety of applications, including natural language processing, question answering machines, and recommendation systems [4]. Companies such as Facebook, Amazon, IBM, LinkedIn, Uber, and Airbnb utilise KGs in their data infrastructure [5, 6].

The Resource Description Framework (RDF) is one of the most common types of KG models [7]. RDF is a standard language for data representation and interchange on the Web. RDF graphs are widely used in practice and adhere to World Wide Web Consortium (W3C) standards [8, 9].

One of the challenges in RDF-based KGs, is representing "time" because RDF itself does not provide a standardised way of representing temporal information. RDF is primarily a framework for representing knowledge in terms of subject-predicate-object triples, and it does not have a built-in way of representing time as a concept [10].

Temporal data modelling in KGs refers to the practice of modelling time-varying information in a structured manner. This allows the KGs to represent not just static facts but also the evolution of entities and relationships over time. By capturing when each triple is valid or occurred, temporal KGs enable more dynamic and accurate modelling of real-world information, reflecting changes, events, or states as they happen. This temporal dimension is essential for applications that require understanding of historical contexts, tracking progression of events, or making time-aware analyses and decisions. Temporal data models ensure time consistency in a data model and assist in analysing historical data, ultimately improving accuracy and decision making [11-14].

Research on modelling temporal data using KGs has seen several RDF-based approaches emerge, including RDF-Star, Named Graph, and the Singleton Property [15-17]. Additionally, the ability to express n-ary relations in RDF has been a significant interest in

the Semantic Web research community since the initial RDF specifications. This capability, commonly referred to as "reification" or "metadata representation," enables the annotation of metadata at the statement level. It is particularly valuable for adding details such as time [15].

There is limited evidence in the area of evaluating temporal data modelling in KGs and the need to conduct more research in this area, have been highlighted in previous literature [15-17]. Firstly, there is no methodology for identifying temporal data modelling approaches in KGs and as such it is difficult to evaluate the comprehensiveness and robustness of the findings. Additionally, the approaches available for modelling temporal data in KGs each come with their own set of benefits and drawbacks. This complicates the process of selecting the most suitable approach for specific needs and highlights the importance of conducting evaluations on these approaches to determine their suitability for specific applications.

This thesis aims to fill the gaps in evaluating temporal data modelling approaches in KGs by proposing a comprehensive metrics framework named TEDME-KG metrics framework. The TEDME-KG metrics framework facilitates the evaluation of temporal data modelling in KGs for Semantic Web researchers. To validate the appropriateness and general usability of the framework, it was assessed by Semantic Web researchers using a survey. Additionally, to validate applicability of the framework, it was assessed by evaluating various temporal data modelling approaches in the context of answering a variety of temporal question categories relating to medication management for patients with a rare disease.

In terms of scope, the research presented in this thesis addresses the design and evaluation of of the metrics framework in terms of appropriateness, general usability, and applicability. The research is not focused on assessing the effectiveness, efficiency, and satisfaction of the framework. In this thesis the metrics framework was applied to a healthcare use case; however, the final metrics framework could potentially be used in other contexts. Moreover, the research does not aim to select the best approach among temporal data modelling approaches in KGs. Instead, the framework has the potential to assist Semantic Web researchers to consider various aspects when modelling temporal data in KGs.

The motivation behind choosing healthcare as a use case is detailed in the following sections.

1.1.1 Healthcare Use case

The healthcare domain is a prime example where the volume and diversity of data are rapidly increasing over time [18]. Healthcare data are produced by a wide range of stakeholders, including physicians, nurses, radiologists, and specialists, and originate from various heterogeneous sources such as hospitals, patient registries, clinics, and laboratories. This data includes time-varying information such as diagnosis time, medication time, visit time, etc., all of which is referred to as temporal data. Temporal data in healthcare are vital for determining prognosis, diagnosis, and therapies. For instance, physicians require detailed clinical histories, encompassing laboratory results, symptoms, past diseases, therapies, medications, and other time-stamped information, to assess a patient's status. Monitoring temporal changes in clinical parameters, such as blood pressure and vital signs, especially in intensive care units (ICUs), is essential for patient care. Health information systems, such as electronic health records (EHRs), personal health records (PHRs), disease registries, and mobile health apps, integrate and manage temporal data to deliver patient care [19, 20].

One specific area in healthcare where temporal data modelling is crucial is the management of rare diseases [21]. Rare diseases often require long-term monitoring and detailed tracking of patient histories, treatments, and outcomes [22]. Temporal data modelling enables healthcare providers to explore disease progression patterns, assess the clinical history of patients, determine potential causes of clinical events over time, and predict future trends based on historical and current data [19, 23].

Medication management in rare diseases also heavily relies on temporal data modelling. Monitoring the intervals and time points of medication consumption is critical for managing treatment efficacy and patient compliance. Accurate temporal modelling of medication start date and stop date helps in understanding the impact of treatments over time, optimising medication plans, and preventing adverse drug interactions [21, 24].

The healthcare use case selected for this research focuses on Anca vasculitis, a type of rare disease. The use case comes from the FAIRVASC project [25, 26]. Bringing together computer scientists, clinicians, and patient organisations, FAIRVASC is a research project

of the European Vasculitis Society and RITA European Reference Network. Ten European partners, representing all facets of patient care for patients with the uncommon illness ANCA vasculitis, make up this partnership. The FAIRVASC ontology [1] which was developed to handle ANCA vasculitis data, is predicated on the standardisation of terminology across these registries.

The research involves modelling the medication start date and stop date of patients, specifically tracking the intervals and time points of medication consumption. This use case is used to demonstrate the practical applicability of the TEDME-KG metrics framework in a real-world setting.

1.2 Research Question

The research question investigated in this thesis is:

"To what extent can a metrics framework facilitate the evaluation of temporal data modelling approaches in knowledge graphs for Semantic Web researchers?"

The terms in the research question can be described as follows:

Metrics Framework: A framework is a logical structure for classifying and organising complex information (ISO/TR 13054: 2012). In this regard, the metrics framework is a set of metrics which are categorised in a logical structure and is used to evaluate temporal data modelling approaches in knowledge graphs.

Facilitate: Make (an action or process) easy or easier [41]. In the context of this research, an action or process is evaluating temporal data modelling approaches in KGs and facilitation is inferred through the following criteria:

- **Appropriateness**: In the context of this research, appropriateness is determined by a set of criteria: relevance, directness, operationality, understandability, comprehensibility, and unambiguity [27] .
- **Applicability**: Extent to which the results are likely to impact on practice [28]. In the context of this research, it focuses on the applicability of TEDME-KG metrics framework on evaluating different temporal data modelling approaches in KGs.

[1] https://ontologies.adaptcentre.ie/fairvasc/

- **Usability:** Usability is defined as the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use (ISO 9241-11:2018). In this research, general usability is assessed without separately evaluating effectiveness, efficiency, and satisfaction.

Temporal data modelling approaches in KGs: Temporal data modelling approaches in KGs are the approaches used to incorporate time-related information into the triples that constitute the graph structure [29].

Semantic Web researchers: Semantic Web researchers, study and develop technologies related to the Semantic Web. Researchers in this field typically work on topics such as ontologies, KGs, semantic data integration, linked data, and semantic technologies that enable more intelligent and meaningful interactions with web content. Their work often involves developing standards, algorithms, frameworks, and applications that leverage semantic technologies to organise, retrieve, and analyse information more effectively [30, 31].

1.3 Research Objectives

In order to address the research question defined above, the following research objectives (RO) were identified.

1.3.1 Research Objective 1 (RO1)

RO1: Identify the problem context and use case motivation by:

- Development of a framework to assess the quality of data sources in healthcare settings and investigate the importance of time and temporal data as a criterion for the quality of data and data sources.
- Exploring the FAIRVASC use case and identifying the associated challenges

1.3.2 Research Objective 2 (RO2)

RO2: Identify the state of the art and gaps in temporal data modelling in KGs by:

- Exploring the approaches for modelling temporal data in KGs

- Exploring approaches for modelling temporal data in healthcare.
- Investigating methods for evaluating data modelling approaches in KGs.

1.3.3 Research Objective 3 (RO3)

- RO3: Design a metrics framework for evaluating temporal data modelling approaches in KGs.

1.3.4 Research Objective 4 (RO4)

- RO4: Evaluate the developed metrics framework in terms of its appropriateness, applicability, and general usability as perceived by Semantic Web researchers.

1.4 Research Methodology

The research methodology employed in this thesis systematically address the research objectives (ROs) through a comprehensive blend of both qualitative and quantitative methods. The methods ensure that each research objective is achieved through a series of structured and iterative processes. The methods utilised in this study include literature reviews for foundational insights, expert advisory group meetings for expert validation, thematic analysis for identifying key themes, surveys for gathering broader perceptions, and experimental study. Furthermore, the Select, Evaluate, and Document approach was used to provide a high-level overview of the TEDME-KG metrics framework development process [27, 32]. During this process, the Goal-Question-Metric (GQM) approach was applied to develop and refine the metrics [33].

The following sections detail how these methods were applied to achieve the specific research objectives and answer the research question outlined in this thesis.

1.4.1 Methodology to achieve RO1

In order to identify the problem context and use case motivation for this research (**RO1**), firstly, a framework was designed to assess the quality of data sources in healthcare settings. This framework highlighted the importance of time and temporal data as a criterion for the

quality of data and data sources. To design the framework, a study was conducted in three phases. In phase 1, a literature review followed by an inductive thematic analysis was conducted to identify items relating to the quality of data sources in healthcare. In phase 2, two expert advisory group meetings were performed to elicit experts' perception of the literature review results and investigate experts' opinions of aspects of quality of data sources in healthcare settings. Finally in phase 3, a framework was developed based on the result of literature review and expert advisory group meetings. The proposed framework is a set of parent themes and sub themes related to the quality of data sources. This study situated healthcare temporal data within a quality framework thereby identifying a context for this research. In the next step, an exploration of the FAIRVASC use case, along with its associated challenges regarding temporal data modelling, was conducted by attending several meetings with the FAIRVASC research group. The studies associated with RO1 are discussed in chapter 2.

1.4.2 Methodology to achieve RO2

To achieve **RO2**, a state-of-the-art literature review was conducted. Firstly, a systematic review was performed to identify temporal data modelling approaches and applications in KGs. Then, a literature review was performed to identify existing literature on temporal data modelling evaluation methods in KGs. As the analysis progressed, it became clear that there is a gap in comprehensively evaluating temporal data modelling approaches in KGs and a well-defined metrics framework may address this gap. In the next step, healthcare temporal data modelling approaches were investigated in depth. A literature review was performed to identify temporal data modelling approaches and applications in healthcare. This study provided a general overview of temporal data modelling in healthcare discussed the gaps in this area. Additionally, methods for evaluating data modelling approaches in KGs were investigated. The studies associated with RO2 are discussed in chapter 3.

1.4.3 Methodology to achieve RO3 and RO4

To achieve RO3 and RO4, a **Te**mporal **D**ata **M**odelling **E**valuation Metrics **Fr**amework in **KG**s (TEDME-KG metrics framework) was designed and evaluated. The framework was designed based on the Select, Evaluate, and Document approach, with overlapping phases to ensure a comprehensive and iterative evaluation process.

The Select, Evaluate, Document approach is a structured methodology used in the metric development process. It was initially proposed by Keeney et al and further improved by

McKay et al. [27, 32]. This approach ensures a systematic and iterative development of metrics, facilitating their validation and applicability.

Select: The Select step involves the identification and selection of relevant metrics. This process typically includes:

- **Literature Review:** Conducting comprehensive reviews of existing literature to identify potential metrics used in similar contexts.
- **Expert Consultation:** Engaging with domain experts to gather insights and recommendations on relevant metrics.
- **Preliminary Selection:** Compiling a list of potential metrics based on their relevance, and alignment with the specific goals of the study or project.

Evaluate: The Evaluate step focuses on the empirical validation and refinement of the selected metrics. This process ensures that the metrics are applicable in real-world scenarios. It includes:

- **Empirical Testing:** Conducting experiments or pilot studies to test the selected metrics in practical settings.
- **Feedback Collection:** Gathering feedback from stakeholders, including experts and end-users, to assess the metrics' usability and relevance.
- **Iterative Refinement:** Continuously improving the metrics based on empirical data and feedback, ensuring they accurately measure the intended outcomes.

Document: The Document step ensures comprehensive and transparent documentation of the entire metric development process. This documentation is crucial for reproducibility, clarity, and future reference. It includes:

- **Detailed Descriptions:** Providing clear definitions and explanations for each metric, including their definition and how they are measured.
- **Methodological Transparency:** Documenting the processes and methodologies used in selecting and evaluating the metrics, including literature reviews, empirical tests, and feedback mechanisms.
- **Guidance for Application:** Offering practical guidelines and examples on how to implement the metrics in real-world scenarios, ensuring their effective use.

By following the Select, Evaluate, Document approach, researchers and practitioners can develop robust and reliable metrics that are well-documented, empirically validated, and tailored to their specific needs. This structured methodology promotes the creation of metrics that are both theoretically sound and practically applicable.

In this research, the Select, Evaluate, and Document approach was employed through four phases in which the TEDME-KG metrics framework was designed and evaluated.

In the **first phase**, a literature review followed by an inductive thematic analysis was conducted to identify metrics for evaluating temporal data modelling approaches in KGs. Then, the first iteration of metrics framework was developed based on the result of literature review. This framework contains different metrics and metrics categories.

In the **second phase**, an experiment was conducted as a pilot study to evaluate two selected temporal data modelling approaches using the preliminary metrics framework. In this step, the framework was updated as a second iteration. The purpose of the second phase is outlined below:

- **Verify Theoretical Assumptions**: To confirm that the theoretical underpinnings of the metrics framework align with actual data modelling practices and outcomes in RDF-based KGs.
- **Test Framework Applicability**: To assess how well the metrics framework can be applied to various temporal data modelling approaches and its effectiveness in providing meaningful evaluations.
- **Refine and Improve**: To identify any shortcomings or areas for improvement in the metrics framework based on real-world testing, leading to further refinements.

During the KG design process, the first iteration of the metrics framework was reviewed using the Goal Question Metrics (GQM) model which was initially proposed by Basili et al [33]. The GQM model is a systematic model for defining and interpreting metrics and involves three levels: (Figure 1)

Goal: At the highest level, for each metrics category, a specific goal was established to provide a clear direction for the evaluation process.
Question: For each goal and its corresponding metrics subcategories, a set of questions was formulated to assess whether the goals are being met. These questions help in translating the abstract goals into concrete queries that can guide the evaluation process.
Metric: For each question, metrics were identified or updated based on the requirements that emerged during the KG design process. These metrics provide the necessary data to answer the questions effectively.

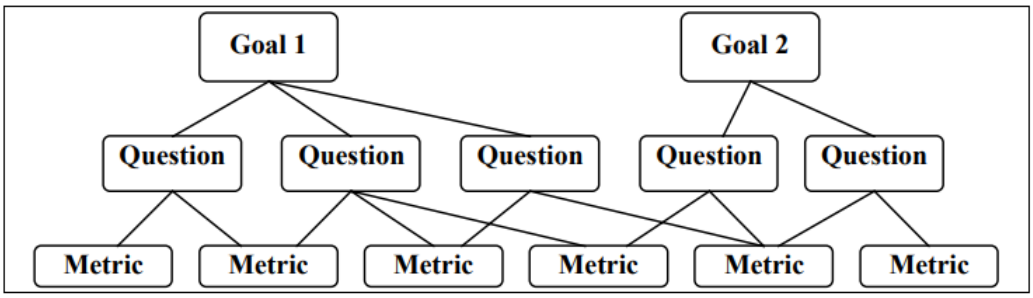


Figure 1- The Goal-Question-Metric approach – (Adapted from Basili et al)

In the **third phase**, a survey was conducted to find out Semantic Web researchers' perception of the framework and the appropriateness, general usability and potential usefulness of the framework was evaluated. The metrics framework was then updated using the experts' feedback and the final iteration of the metrics framework was developed.

In the **fourth phase,** an experimental study was conducted, and the applicability of the framework was evaluated using a real-world scenario and different temporal questions. Throughout the 4 phases, the metrics were documented using three headings including definition, unit, possible ranges/answers.

The design and evaluation of the TEDME-KG metrics framework fulfilled **RO3 and RO4** and are discussed across three chapters. Chapter 4 covers the design, explaining phases one and two. Chapter 5 focuses on the evaluation by experts, detailing phase three. Finally, Chapter 6 presents the experimental study, which explains phase four.

1.5 Thesis contribution

The primary contribution of this thesis is the development of the Temporal Data Modelling Evaluation Metrics Framework in Knowledge Graphs (TEDME-KG metrics framework). Additionally, the thesis offers several minor contributions: a framework for assessing the quality of data sources in healthcare, a categorised list of temporal data modelling approaches and their applications, and the evaluation results from the experimental study conducted using the healthcare use case and the TEDME-KG metrics framework.

1.5.1 Major contribution

The major contribution of this thesis is the Temporal Data Modelling Evaluation Metrics Framework in KGs (TEDME-KG metrics framework). This framework is designed to evaluate temporal data modelling approaches within RDF-based KGs, providing a

comprehensive tool for assessing these approaches in handling temporal data. The framework consists of 3 goals, 3 metric categories,7 metrics subcategories, and 16 individual metrics, along with questions related to each goal. Definitions, units, and possible ranges for each metric are also provided.

1.5.2 Minor contribution 1

The first minor contribution of this thesis is the proposed framework to assess the quality of data sources in the healthcare settings. The framework includes eight parent themes and 22 sub themes. The proposed framework can assist people using different data sources in healthcare to identify and assess various aspects of the quality of a data source. It can also guide policy makers and health information standard organisations to establish standards, guidelines, legislation regarding data sources in healthcare settings considering the themes proposed in this framework. Moreover, this framework provides guidance for health data sources development team to consider different aspects of a data source prior to the design and implementation of a data source.

1.5.3 Minor contribution 2

The second minor contribution is a categorised list of approaches and applications of temporal data modelling in KGs. This list was compiled through a systematic review process and provides a comprehensive overview of the existing approaches and their respective use cases, offering valuable insights for researchers and practitioners in the field.

1.5.4 Minor contribution 3

The final minor contribution includes the evaluation results derived from the experimental study conducted using the proposed TEDME-KG metrics framework. These experiments validate the framework's applicability in evaluating different temporal data modelling approaches, providing empirical evidence to support the framework's applicability in real-world applications.

1.6 Impact

The TEDME-KG metrics framework, and categorised list of temporal data modelling approaches in KGs collectively enhance the field of temporal data modelling in KGs. One of the most impactful aspects of this thesis is its application in a real-world case study in the healthcare domain. The TEDME-KG metrics framework has been instrumental in evaluating different temporal data modelling approaches, which in turn, has guided the extension and

improvement of the FAIRVASC ontology which was used as a use case in this research. During the design of different KGs based on different temporal data modelling approaches, new data modelling solutions were proposed to the FAIRVASC team for modelling medication times. There were further discussions with the FAIRVASC team regarding the application of temporal data modelling approaches to other patient-related time-sensitive data, such as biomarkers that fluctuate over time.

The data source quality assessment framework was published in a prestigious Journal of the American Medical Informatics Association (JAMIA). It was also presented at the National Patient Safety Office Conference and annual Health Information and Quality Authority (HIQA) PhD showcase event for all the HIQA staff. Additionally, Science Foundation Ireland (SFI) featured a post on their website and conducted an interview with the researchers involved in this study.

1.7 Publications

This publications section is divided into published/accepted papers and publication plan.

1.7.1 Published papers.

The published/accepted/presented papers associated with the research in this thesis to date are:

1. Hooshafza S, Mc Quaid L, Stephens G, Flynn R, O'Connor L. **Development of a framework to assess the quality of data sources in healthcare settings**. J Am Med Inform Assoc. 2022 Apr 13;29(5):944-952.

   This paper, associated with RO1, highlights the importance of time and temporal data as critical criteria for assessing the quality of data and data sources. It proposes a comprehensive framework for evaluating various data sources within healthcare settings. The paper was published in the prestigious Journal of the American Medical Informatics Association (JAMIA). Additionally, Science Foundation Ireland (SFI) featured a post on their website and conducted an interview with the researchers involved in this study.

2. Hooshafza S, Orlandi F, Flynn R, Mc Quaid L, Stephens G, O'Connor L. **Modelling temporal data in knowledge graphs: a systematic review protocol**. HRB Open Res 2022, 4:101

   This paper, associated with RO2, presents a systematic review protocol aimed at identifying and evaluating temporal data modelling approaches in KGs. It provides a detailed methodology for conducting the review, including criteria for selecting studies, data extraction processes, and methods for synthesising findings. The goal of the systematic review was to establish a comprehensive understanding of current approaches in temporal data modelling within KGs. This research was published in the Health Research Board (HRB) Open, an open-access journal, ensuring wide accessibility and dissemination of the findings to the global research community.

3. Hooshafza S, Mc Quaid L, Stephens G, Flynn R, O'Connor L. **Enhance quality health systems by quality data sources**, National Patient Safety Office Conference, October 2022, Dublin, Ireland.

   This paper, associated with RO1, discusses the importance of a data source quality assessment framework as a tool for evaluating data sources in healthcare settings. It highlights the need for thorough assessment of data sources in healthcare, considering various aspects. This research underscores the significance of robust data sources quality frameworks in enhancing patient safety and healthcare outcomes. The paper was presented at the National Safety Office Conference.

4. Hooshafza S, Stephens G, Little M, Yaman B. **Temporal Data Modelling Evaluation in Knowledge Graphs: A Healthcare Use Case**, SWAT4HCLS 2024: The 15th International Conference on Semantic Web Applications and Tools for Health Care and Life Science, February 26–29, 2024, Leiden, Netherlands

   This paper, associated with RO3, details the evaluation process of various temporal data modelling approaches. It presents the results of these evaluations within the context of a healthcare use case, highlighting the applicability of two selected approaches. This research was presented at the 15th International Conference on Semantic Web Applications and Tools for Healthcare and Life Sciences.

5. Hooshafza S, Stephens G, Randles A, Little M, and Yaman B. **TEDME_KG Metrics Framework: A Metrics Framework for Temporal Data Modelling Evaluation in Knowledge Graphs**. KGSWC 2024: Sixth International Conference in Knowledge graph and Semantic Web, December 06–08, 2024, Paris, France

   This paper, associated with RO3, details the design process of the TEDME-KG metrics framework. It begins with a comprehensive literature review aimed at identifying and extracting relevant metrics. Following this review, the paper describes an experiment conducted to update and refine the metrics framework, ensuring its robustness and applicability. The results of this study underscore the framework's potential in enhancing KG and Semantic Web technologies. This research has been presented at the Sixth International Conference in Knowledge Graph and Semantic Web (KGSWC 2024).

1.7.2 Publication Plan

The publication plan associated with the research in this thesis is:

1. Hooshafza S, Stephens G, Randles A, Little M, and Yaman B. **Application of temporal data modelling in knowledge graphs in a rare kidney disease registry**. Health Informatics Journal.

   This paper, associated with RO3, presents a comprehensive evaluation of various temporal data modelling approaches within the context of a healthcare case study. The study provides detailed analysis and comparison of the selected temporal data modelling approaches, highlighting their applicability in different aspects. The insights gained from this evaluation are crucial for advancing data management practices in medical informatics. This paper has been prepared for submission at the Health Informatics Journal.

2. Hooshafza S, Stephens G, Randles A, Little M, and Yaman B. **Evaluating the appropriateness and general usability of a metrics Framework for Temporal Data Modelling Evaluation in Knowledge Graphs**. Semantic Web Journal.

This paper, associated with RO4, presents the results of evaluating the appropriateness and general usability of the TEDME-KG metrics framework. The study offers an in-depth analysis of the framework's appropriateness and general usability in practical applications. The findings highlight its potential benefits and areas for improvement, contributing valuable insights to the field. This paper has been prepared for submission at the Semantic Web Journal.

1.8 Thesis Overview

The remainder of this thesis is structured as follows.

1.8.1 Chapter 2 Problem Context Identification

This chapter addresses the problem context and use case motivation. It begins with the development of a framework to assess the quality of data sources in healthcare settings, highlighting the importance of time and temporal data as criteria for data quality. Additionally, it delves into the FAIRVASC use case and disuses the associated challenges. These efforts collectively establish a context for understanding the significance of temporal data in healthcare and the necessity of robust evaluation tools.

1.8.2 Chapter 3 State of the Art

This chapter provides an in-depth review of the current state of the art in modelling temporal data using knowledge graphs and in healthcare. It includes a review of various approaches for modelling temporal data in knowledge graphs and investigates different methods for evaluating these approaches. Additionally, it reviews temporal data modelling approaches in healthcare. The purpose of this chapter is to provide a comprehensive overview of existing approaches, identify gaps, and highlight opportunities for further research in temporal data modelling. This chapter sets the stage for understanding the advancements and challenges in the field, providing a solid foundation for the research presented in the subsequent chapters.

1.8.3 Chapter 4 Design

This chapter outlines the development of the TEDME-KG metrics framework, following the Select-Evaluate-Document process discussed in section 1.4. The initial phase involved a

literature review and an inductive thematic analysis to pinpoint metrics suitable for evaluating temporal data modelling approaches in Knowledge Graphs. This led to the creation of the first iteration of the metrics framework, which integrates a variety of metrics and categories.

The second phase featured a pilot study of two chosen temporal data modelling approaches using this initial framework. During this phase, the Goal-Question-Metric (GQM) approach was utilised to refine and update the framework, resulting in its second iteration.

1.8.4 Chapter 5 Evaluation

This chapter discusses the evaluation of the TEDME-KG metrics framework. It involved conducting a survey to collect Semantic Web researchers' opinions on the framework, specifically examining its design, appropriateness, and general usability. Feedback from these experts was then incorporated to refine and finalise the metrics framework.

1.8.5 Chapter 6 Implementation

This chapter presents an experimental study that employed a real-world scenario to evaluate various temporal data modelling approaches through different temporal queries. The final version of the TEDME-KG metrics framework was implemented using the FAIRVASC use case. The study aimed to evaluate the framework's applicability in practical settings.

1.8.6 Chapter 7 Conclusion

This chapter provides a comprehensive summary of the research findings. It discusses how the research question and objectives have been addressed, highlights the key contributions of the thesis, including the development and evaluation of the TEDME-KG metrics framework, and explores the implications of the findings for knowledge graphs, temporal data modelling, and healthcare data management. The chapter also outlines the limitations of the research, offers recommendations for future studies, and suggests potential avenues for further exploration, emphasizing the broader impact of the work on both theoretical and practical aspects of data modelling in healthcare.

Chapter 2 Problem Context Identification

This chapter discusses the problem context and the motivation behind the use case. It addresses Research Objective 1, which is intended to answer the corresponding research question. A framework for evaluating the quality of data sources within the healthcare domain was developed, highlighting the role that time and temporal data play in assessing both data and data source quality in healthcare. The chapter further discusses the FAIRVASC use case which is the used in this thesis as a real-world experimental study. This use case related to the healthcare domain and focus on temporal data modelling.

Together, these studies underscore the importance of temporal data in healthcare and the necessity for temporal data modelling.

2.1 Data source quality assessment framework

This section discusses the process involved in developing the data source quality assessment framework.

2.1.1 Background in data source quality assessment in healthcare

The rapid evolution of accessible electronic healthcare data underpins a wide variety of activities including research, clinical management of primary care, health system planning, documentation of clinical care, and epidemiological surveillance. These activities provide many opportunities for making improvements in healthcare delivery, healthcare decision making, and patient care [34].

There are a wide variety of electronic healthcare data sources including electronic health records (EHRs), laboratory systems, imaging systems, patient and/or disease registries, medical devices, remote monitoring devices and wearable devices [35]. Healthcare data are generated as part of service provision to patients, or through disease monitoring. There are also patient generated health data which are recorded, created or gathered by patients (or family members) through apps and wearables [36]. Healthcare data are used by different health information users including service users and patients to make informed decision about their own health, clinical staff to provide appropriate care for patients, administrative

staff to arrange outpatient clinics, social care staff to provide health services, government departments to develop health and social care policy and researchers to analyses and interpret data, and make conclusions [37].

Providing access to and use of high-quality data are key to ensuring safe, reliable health and social care. High quality data are essential not only for providing high quality clinical and social care and continuity of care, but also for decision making, strategic planning and management within an organisation. Poor collection and management of information has negative impacts on the health and social care system [37].

In this regard, data quality frameworks indicate approaches and tools to systematically assess, document and improve data quality within an organisation [38]. This has been highlighted in previous studies which focus on developing data quality frameworks for different purposes including EHR data, internet of things (IOT) devices, cloud based systems [39-41]. In 2012, a data quality guide was published in Ireland by the Health Information and Quality Authority (HIQA) for health and social care staff. It discussed seven essentials for improving data quality to support health and social care and subsequently the overall quality of care. This guide considered "data quality dimensions" as one of the essentials. The essentials proposed in the guideline are, leadership and management, policies and procedures, standardisation, training, data quality dimensions, performing data quality audits, and making data available [37]. Moreover, international frameworks exist regarding data quality for example, the Canadian Institute for Health Information (CIHI)'s data quality framework [42], Australian Capital Territory (ACT) health data quality framework and Australian Bureau of Statistics (ABS) data quality framework [43, 44], and HIQA guidance on data quality framework for health and social care in Ireland [38].

Thus far, the focus has been mostly on data quality and there has been no systematic synthesis of studies concerning data source quality. In the healthcare context, a data source refers to any origin or system that provides health-related data.

Using health data sources without considering quality evaluations can lead to negative consequences in the health domain, including poor interpretation and meaningfulness of the final results and outcomes. Therefore, it is crucial to improve the quality of data sources in combination with the quality of data [45].

The need to perform more research and develop tools to assess the quality of data sources has also been highlighted in a user's guide published in 2019 by the Agency for Healthcare Research and Quality [36].

Preliminary searches of the existing literature, identified two documents regarding the quality of data sources in healthcare settings, the CIHI's data source assessment tool and the Australian Institute of Health and Welfare (AIHW) data source assessment framework [46, 47]. However, the methodologies for both have not been described and as such it is difficult to evaluate their comprehensiveness and robustness. The CIHI's data source assessment tool was published in 2017 [46]. The tool contains two sets of criteria. A general set for assessing the quality of all types of data sources and a supplementary set for assessing the quality of survey results or sample data. Five dimensions of quality including relevance, accuracy and reliability, comparability and coherence, timeliness and punctuality, accessibility and clarity, were considered to shape the CIHI's tool. The AIHW's framework for assessing data sources for population health monitoring was published in 2014 [47]. This framework contains a three-step process which includes collecting details of data sources, identifying the potential to inform key monitoring areas, and assessing the quality of data to specify its fitness for purpose.

2.1.2 Purpose

Based on existing literature, the quality of data sources in healthcare has been insufficiently explored, and a more systematic and comprehensive analysis is needed to identify various aspects related to the quality of data sources in healthcare. In this study, an assessment framework was developed to evaluate the quality of data sources in healthcare. The goal was to first identify existing data source quality items from the literature, combine them with expert opinions on the quality of data sources, and ultimately develop a framework to assess their quality in healthcare settings.

To better understand the concept of data source quality, it's essential to differentiate it from data quality. Data quality refers to the inherent characteristics of the data itself, such as accuracy, completeness, consistency, timeliness, and reliability. It focuses on evaluating how well the data meets its intended purpose or use. In other words, data quality examines the attributes of the data that determine its fitness for decision-making or analysis. However, while data quality assesses the data's fitness for purpose, it does not address the underlying origin or context of the data.

In contrast, data source quality refers to the quality of the sources from which the data is obtained. This involves factors such as the reliability of the data source, its credibility, the processes used to collect and manage the data, and its potential biases. Data source quality concerns the trustworthiness and the overall framework surrounding the data source, which directly impacts the quality of the data it produces. Therefore, while data quality focuses primarily on the attributes of the data, data source quality takes a broader perspective, encompassing the governance, management, and origin of the data.

There is, however, significant overlap between these two concepts. The quality of a data source can profoundly influence the quality of the data it provides. For instance, if a data source is unreliable or poorly managed, the data collected from that source is more likely to be inaccurate or inconsistent. Nevertheless, data quality tends to concentrate more on the specific characteristics of the data, while data source quality includes considerations about the data source's integrity and operational context.

This distinction is vital because evaluating data source quality provides insights into the potential risks or limitations inherent in the data, offering a more holistic view of the data's origins. In comparison, data quality focuses on assessing how well the data serves its intended function in a particular use case. By systematically synthesizing data source quality, this study aims to contribute to the field, enabling a more comprehensive understanding of the broader factors that influence data quality and, ultimately, improving the reliability of data in healthcare settings.

2.1.3 Methods

This study was conducted in three phases. In the first phase, a literature review was performed to identify and extract items relating to the quality of data sources in healthcare. In the second phase, HIQA expert advisory group meetings were held to find out experts' perception of the results of the literature review and to investigate experts' opinions in relation to the aspects of quality of data sources in healthcare settings. In the third phase, a framework was developed based on the literature review and the advisory group meeting results. (Figure 2)

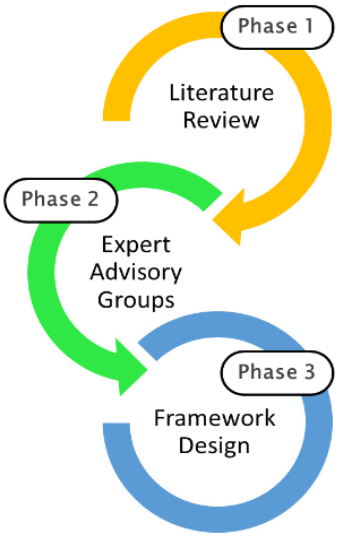


Figure 2- Development phases of data source quality assessment framework in healthcare settings

2.1.4 Phase 1: Literature review

2.1.4.1 Formulating the research question

The literature review was conducted using the CIMO (Context, Intervention, Mechanism, and Outcome) question framework for developing the research question as follows, Context: Healthcare data sources, Intervention: Assessment, Mechanisms: Framework, and Outcome: Quality [48].

The research question was, what are the assessment criteria related to the quality of data sources in healthcare settings?

2.1.4.2 Information sources

Searches was carried out on the following databases:

- Library and Information Science Abstracts (LISA) accessible by ProQuest
- Library, Information Science and Technology Abstracts accessible by EBSCOhost
- MEDLINE accessible by EBSCOhost
- IEEE Xplore

2.1.4.3 Search strategy.

The search concepts and keywords based on CIMO framework are set out are set out in Table 1.

Table 1- Search concepts and keywords for use in a literature review to identify areas of quality relating to data sources in healthcare settings

CIMO framework	Concepts	Keywords
Context	Healthcare data sources	("Data source" OR "Information source" OR "knowledge source" OR Repositor* OR Registr* OR "Data file" OR "Data feed" OR "Data origin") AND (healthcare or "health care")
Intervention	Assessment	Evaluat* OR Assess* OR assurance OR critique
Mechanisms	Criteria	Framework OR Tool OR Checklist OR Criteria OR Mechanism
Outcome	Quality	Quality OR Appropriateness OR suitability OR trust

The search query used for Library and Information Science Abstracts (LISA), as an example of the search strategy, was:

(("Data source" OR "Information source" OR "knowledge source" OR Repositor* OR Registr* OR "Data file" OR "Data feed" OR "Data origin") AND (healthcare or "health care") AND (Evaluat* OR Assess* OR assurance OR critique) AND (Framework OR Tool OR Checklist OR Criteria OR Mechanism) AND (Quality OR Appropriateness OR suitability OR trust))

The terms 'data' and 'dataset' were not included as standalone keywords because they are overly broad and tend to retrieve a large volume of irrelevant or generic results. Instead, the term "data source" and its variants (e.g., "information source," "repository," "register," "data feed") were used to narrow the scope specifically to sources from which data is obtained, which aligns better with the focus of this study.

Similarly, 'metric' and 'dimension' were not explicitly included because the aim was to explore frameworks, tools, and criteria used for assessment, rather than to capture every individual measurement term. The terms "framework," "tool," "checklist," and "criteria" were selected as they reflect structured approaches that inherently include metrics and dimensions.

A search for grey literature including national or organisational reports, standards, and guidelines was also performed, in Google, using the same search concepts.

2.1.4.4 Inclusion and exclusion criteria

The search criteria were limited to literature that were published in English between January 1, 2010 and 29 September 2021. To be eligible for inclusion in the review, a criterion was that the publications needed to contain information on data source quality assessment frameworks in a healthcare setting. We excluded studies that focused solely on data quality, opinion/review papers, and qualitative perception papers.

2.1.4.5 Screening

Retrieved literature were imported into Covidence, systematic review software by Veritas Health Innovation Ltd, for screening [49]. Titles and abstracts were screened for inclusion by one researcher. Titles, introductions and contents pages of retrieved grey literature were screened for inclusion by one researcher. Full text review of screened literature was then completed by the same researcher to identify the final list of literature for inclusion.

Screening Limitation

A key limitation of the screening process was that it was conducted by a single reviewer, which increased the risk of subjective bias. To mitigate this risk, the inclusion and exclusion criteria were clearly defined and consistently applied throughout the process. A systematic approach was followed, using specific keywords aligned with the study's focus, which reduced the potential for inconsistent decisions. Additionally, the inclusion criteria and full text of selected papers were cross-checked to ensure relevance and alignment with the research question.

Although the process was conducted by a single reviewer, efforts were made to minimize bias by documenting all decisions made at each stage of the review. This ensured transparency and provided an opportunity for validation if needed.

2.1.4.6 Selection of studies

Search queries returned 4908 papers, of which 36 duplicate papers were removed. Screening of titles and abstracts led to the exclusion of 4793 papers and retention of 79 papers for full text review. After considering the inclusion and exclusion criteria, 8 papers were included

in the literature review. Two grey literature papers were also identified for inclusion. Ten papers were included in the review. (Figure 3)

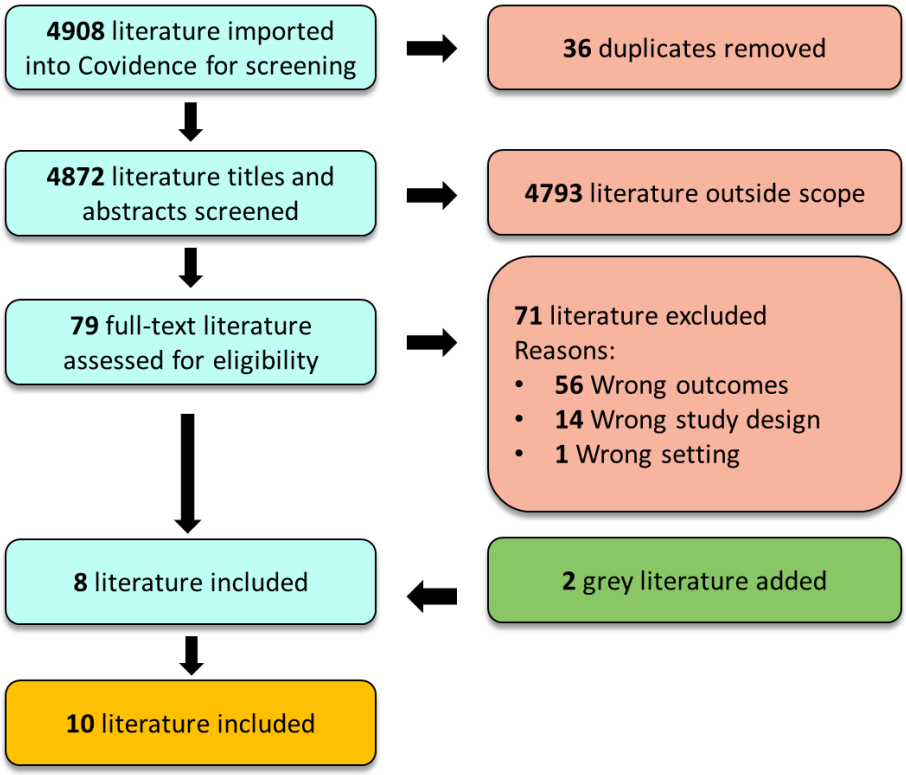


Figure 3-Flow chart of literature screening for review of assessment criteria relating to the quality of data sources in healthcare

2.1.4.7 Data extraction

A data extraction form built in Covidence was used to record information from each included paper. One researcher read the papers in full and manually extracted the data directly into the data extraction table. The reference, setting, and all mentions of data source quality items, were recorded. The data extraction form was then exported into Excel, version 15.0 by Microsoft Corporation as tables, in order to perform analysis [50].

2.1.4.8 Data synthesis

An inductive thematic analysis of the retrieved literature was conducted to identify themes relating to the quality of data sources in healthcare. The thematic analysis was conducted in four steps:

- Step 1: A sub theme was assigned to each extracted item based on their concept. The sub themes were not unique per item. For example, where two items referred to the same concept, the sub theme was repeated for both.
- Step 2: Similar sub themes were grouped together
- Step 3: A parent theme was defined and assigned for each set of sub themes
- Step 4: The themes were reviewed and agreed by a second researcher.

2.1.5 Phase 2: Expert advisory groups

Two HIQA expert advisory group meetings were conducted to capture experts' thoughts on the findings of the literature review and to collect expert input on aspects of quality of data sources and gaps in the findings of the review. Experts were invited to participate from within HIQA, the organisation in Ireland responsible for setting recommendations, standards and guidance on health information, and assessing compliance with those national standards. Seven experts participated. These were from across disciplines (health information, technical standards and management of a national healthcare data collection). To prevent the Halo effect i.e. the cognitive bias that might be applied by the managers to other staff, one meeting was performed with 3 managers and the second meeting was held with 5 project leads. The meeting was facilitated by one researcher. Participation was voluntary and by direct invitation. As per HIQA operational procedures, written consent for participation and use of findings in research outputs was sought and received by email. In order to inform the participants, an information package including a brief summary of the literature review, the identified themes and detail of the advisory group meetings, was sent to participants prior to the meetings. The following overarching questions were asked at the internal expert advisory groups:

1. Are there any other themes that should be added?
2. Are there any themes that should be removed?
3. Do you have any comments on how the sub-themes are grouped?
4. Do you have any comments on the naming of the themes or sub-themes?

Participants were also given time to make any additional comments regarding the themes. Notes on participant's opinions and on the discussion were taken by a note taker and then confirmed by the facilitator. No personally identifiable information was collected, and no effort was made to link what the participants said in any way after the meeting.

To synthesis the results of the advisory groups, a list of comments was extracted from the notes and similar comments were grouped together into categories.

2.1.6 Phase 3: Framework design

The themes extracted from the literature review were structured to form a preliminary framework. Changes were applied based on the results of the advisory group meetings; new themes were added, some themes were removed, categorisation of some themes were changed and, the names of some themes revised. The resultant themes were cross-checked by a second researcher and the themes finalised. A visual model of the final framework was designed.

2.1.7 Results

2.1.7.1 Phase 1: Literature review

Ten papers published between 2012 and 2019 were identified for inclusion. A total of 129 items were extracted including, survey items, strategies, policies, recommendation, process, attributes, capabilities, and characteristics relating to the quality of data sources. (Appendix A)

Themes identification

Eight parent themes and 24 sub-themes were identified. (Figure 4)

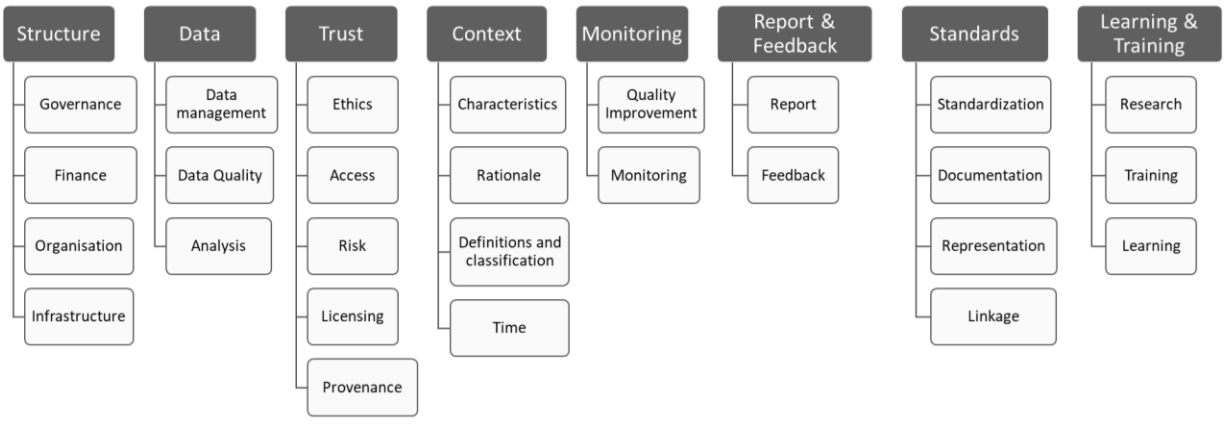


Figure 4- Themes identified in the literature review on data source quality assessment criteria in healthcare

2.1.7.2 Phase 2: Expert advisory groups

The comments of the experts in advisory group meetings were categorised into three different categories, outlined below:

1. Naming themes: New names were proposed for the two parent themes as follows:
    - "Structure" to be replaced by "Governance, leadership and management".
    - "Standards" to be replaced by "standardisation" and the sub theme "standardisation" to be replaced by "standards".
2. Adding or removing themes: New themes and removal of other themes were proposed as follows:
    - Addition of new themes: "security", "dissemination" as sub themes and "use of information" as a parent theme.
    - Removal of themes: "Risk", and "rationale "to be removed because they are covered by "governance" sub theme. Also "report" to be removed since it is covered by "dissemination".
3. Grouping themes: Regrouping of some themes were suggested.
    - "Infrastructure" to be placed under "standardisation".
    - "Analysis", "research", "dissemination" to be placed under "use of information".
    - "Definitions and classifications" to be placed under "standardisation".
    - The sub themes "Monitoring" and "feedback" to be grouped together and called "Monitoring and feedback".

All the above comments were accepted and were then applied to the preliminary framework.

2.1.7.3 Phase 3: Framework design

Synthesis of literature review results and expert advisory groups input resulted in eight parent themes and 22 sub themes. The parent themes were 1. Governance, leadership and management, 2. Data, 3. Trust, 4. Context, 5. Monitoring, 6. Use of information, 7. Standardisation, 8. Learning and training. The 22 sub themes were governance, finance, organisation, characteristics, time, data management, data quality, ethics, access, security, quality improvement, monitoring and feedback, dissemination, analysis, research, standards, linkage, infrastructure, documentation, definitions and classifications, learning and, training. These themes were used to construct a visual model of a framework for data source quality assessment in healthcare. (Figure 5) The definition of each sub theme is provided. The

definitions are based on criteria identified from the literature and subsequently grouped into sub-themes. (Table 2)

Table 2- Data source quality assessment framework - Sub themes definitions

Parent Theme	Sub-Theme	Definition
Governance, Leadership, Management	Governance	Structures and processes for decision-making, oversight, and accountability in data source development and maintenance.
	Finance	Availability and management of financial resources to support data source operation, maintenance, and improvement.
	Organisation	The roles, responsibilities, and institutional arrangements of teams or entities managing the data source.
Context	Characteristics	The inherent features of the data source, including its scope, origin, and purpose.
	Time	Temporal aspects of the data source, such as the time period covered, update frequency, and timestamping practices.
Data	Data Quality	The degree to which data is accurate, complete, consistent, reliable, and relevant for its intended use.
	Data Management	Practices and infrastructure for data storage, handling, integration, and versioning.
Trust	Ethics	Adherence to ethical principles, including consent, fairness, and responsible use of data.

	Privacy	Measures to protect individual privacy, including data anonymization and access control policies.
	Access	The availability and ease with which users can retrieve and use the data, subject to legal and ethical constraints.
Monitoring	Quality Improvement	Systematic activities to improve data processes and outcomes based on performance evaluation.
	Monitoring & Feedback	Ongoing assessment of data source performance and mechanisms for collecting and incorporating stakeholder feedback.
Use of Information	Dissemination	The processes and strategies for sharing data and findings with stakeholders or the public.
	Analysis	The ability and tools available to process and interpret data to generate insights.
	Research	Use of data to support scientific inquiry, hypothesis testing, and evidence-based decision-making.
Standardisation	Standards	Adoption of agreed-upon data formats, terminologies, and protocols to ensure consistency and interoperability.
	Linkage	The capability to connect datasets through common identifiers or semantic relationships.
	Infrastructure	The technical systems and tools supporting data storage, access, analysis, and sharing.

	Documentation	Availability of descriptive metadata, data dictionaries, and procedural records to support transparency and reuse.
	Definitions & Classification	Use of clear, standardized definitions and classification systems to maintain consistency and clarity in data.
Learning and Training	Learning	Processes to promote shared understanding and continuous learning among stakeholders involved in data use or management.
	Training	Provision of training for data contributors, users, and custodians to ensure proper handling and interpretation of the data.

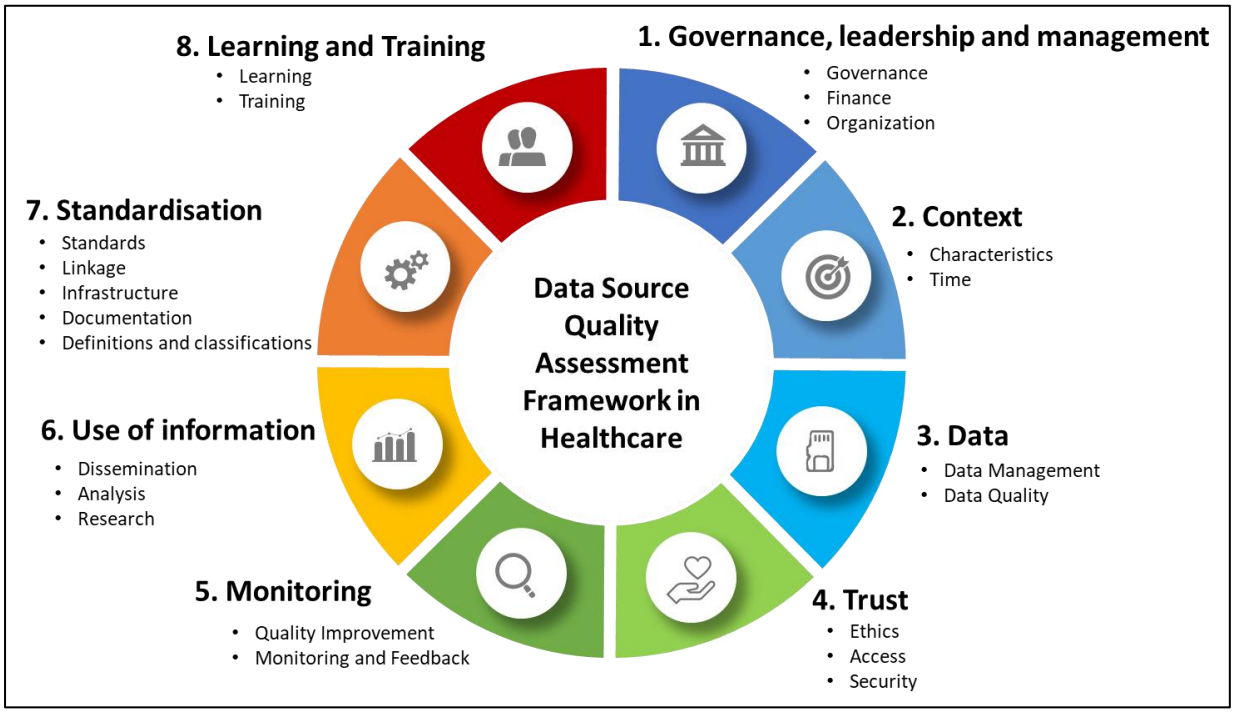


Figure 5- Data source quality assessment framework in healthcare settings

2.1.8 Discussion and conclusion

In this study, a framework was developed for evaluating data source quality in the context of healthcare. This framework was developed using a robust methodology which included a

comprehensive review of the literature to extract previously identified data source quality items, followed by filtering and matching items, developing a list of themes and revising the themes based on expert opinion.

The themes that were identified in the literature included eight parent themes and 27 sub themes. As this is a relatively novel area, there was a risk that the themes in the literature would not be comprehensive, and the terminology used in the literature might not be representative of what is used in current practice. Furthermore, the manner in which sub-themes were grouped, albeit inductive, was also open to subjectivity, by the nature of the thematic analysis. As such, we complimented the findings from the literature review and thematic analysis with input from people with expert working knowledge of health information and data sources in order to improve applicability and reduce subjectivity.

We found no previously published framework for assessing the quality of data sources in healthcare among journal articles. Two grey literature papers proposed frameworks or tools for assessing data sources in healthcare were identified; however, no evidence was found regarding their methodologies. Furthermore, many aspects of quality identified in the literature were not included in these frameworks. The final framework included eight unique parent themes and 22 sub themes related to the quality of data sources in healthcare. Some, but not all, of the themes identified have previously been identified in other similar work.

1. Governance, leadership and management: This parent theme includes governance, finance and organisation. The parent theme was also highlighted as a theme in the Information management standards for national health and social care data collections published by HIQA in 2017; however, in this work, there was no sub category mentioned for this theme and financial aspect had not been addressed under the governance, leadership and management [51]. In our proposed framework, governance, finance and organisation were considered as separate sub themes to highlight the importance of each area.
2. Context: The parent theme context includes characteristics and time related aspect of a data source. In the AIHW framework, the first step towards data source assessment was mentioned “data collection” which included data characteristics and time related aspect of a data source. This is similar to what we have proposed in the herein framework [47]. The difference is that, we defined a parent theme context to cover characteristics and time related aspects of a data source. Moreover, since time was mentioned in many studies, we considered it as a separate sub theme.

3. Data: The parent theme data proposed in this study includes data quality and data management sub themes. We considered both data management and data quality aspects of a data source under one parent theme, to cover all data related issues in terms of quality dimensions, quality characteristics, data elements and storage issues. This consolidation should assist data source users/developers in assessing the quality of the data as data quality and data management are intrinsically linked. In other similar work, the two sub themes have not been grouped [47]. Regarding the data quality sub theme, existent data quality frameworks which propose dimensions, characteristics and criteria for assessing the quality of data, are suggested [38, 42-44].
4. Trust*:* The parent theme trust covers ethics, access and security. In the HIQA Information management standards for national health and social care data collections, the theme was not addressed as a separate category and the sub themes ethics, privacy and access were discussed under different categories [51]. In comparison to the European Commission's Trustworthy AI Guidelines, which conceptualize trust through principles of lawfulness, ethics, and robustness, our data source quality framework operationalizes trust by focusing on ethics, privacy, and access. While the EU guidelines offer a high-level view suited to AI system governance, our model situates trust directly within the management of data sources, ensuring these critical areas are treated interdependently within healthcare data collections [52]. We considered trust as a parent theme to highlight the importance of this area and grouped ethics, privacy and access under it to ensure they are considered in relation to one another.
5. Monitoring*:* The parent theme monitoring proposed in this study includes quality improvement and monitoring and feedback. We propose quality improvement activities of a data source with the focus on quality of process involved in a data source. In addition, we included monitoring information area and providing feedback about different aspects of a data source under this parent theme. Monitoring was also mentioned in the HIQA Information management standards for national health and social care data collections within the proposed standards in this document. It was used for different purposes and mostly related to data quality monitoring [51].
6. Use of information: This parent theme includes dissemination, analysis and research. The parent theme was highlighted in the HIQA Information management standards for national health and social care data collections; however, in this document, the sub themes analysis and research had not been mentioned under the use of information [51]. Other published frameworks have not mentioned use of information [46, 47].

7. Standardisation: This parent theme includes standards, linkage, infrastructure, documentation and definitions and classification of a data source. In our framework, we covered different aspects related to standardisation under one parent theme. This theme addresses all aspects of setting and implementing standards in a data source and can assist data source developers when implementing standards in a data source. This grouping for standardisation, was not proposed in any previous similar works [46, 47, 51].
8. Learning and training: This parent theme covers learning and training aspects of a data source. This theme was not addressed in either the AIHW framework for assessing data sources for population health monitoring or CIHI's data source assessment tool [46, 47]. In our proposed framework, learning and training encompass learning and collaboration between data sources and the provision of training for those submitting data and for data abstractors.

The proposed framework represents the combination of different aspects related to the quality of data sources in healthcare settings. It is designed to be as comprehensive as possible. However, as this is an evolving area, it is likely that updates will be required over time. It must also be noted that, it is unlikely that every data source will necessitate covering all of the aspects, and it can depend on the functionality of a data source.

To the best of the authors' knowledge, this study is the first to use systematic and thematic analysis to identify quality aspects related to data sources in healthcare. The herein publication of the methodology will facilitate future replication and updates to the framework; endeavours that are necessary in this rapidly evolving field.

In terms of limitations, the final included journal articles were largely based on registries and there were no journal articles about other health data sources. This may impact the scope of our proposed framework which aims to be an assessment framework for health data sources in general. However, this issue has been ameliorated by incorporating expert opinion and experience regarding health data sources in general.

The proposed framework can assist people using different data sources in healthcare to identify and assess various aspects of the quality of a data source. It can also guide policy makers and health information standard organisations to establish standards, guidelines, legislation regarding data sources in healthcare settings considering the themes proposed in this study. Moreover, this framework provides guidance for health data sources

development team to consider different aspects of a data source prior to the design and implementation of a data source.

While this framework was developed in the context of healthcare data sources, many of its quality themes—such as standardisation, trust, governance, and data quality—also align closely with established Linked Data quality dimensions. Linked Data quality frameworks typically emphasize dimensions such as accuracy, completeness, interoperability, provenance, timeliness, and accessibility, all of which are reflected either directly or indirectly within the proposed themes. For example, the standardisation theme in our framework supports interoperability through linkage, classification, and infrastructure; the trust theme relates to data access and ethical usage, which are central to Linked Data governance; and the context and data themes address completeness, consistency, and timeliness. While the framework was tailored for healthcare, these overlaps highlight its broader relevance and potential alignment with Linked Data quality principles, providing a foundation for assessing both traditional and semantic web-oriented data sources.

2.2 Time and temporal data in data and data source quality assessment

One of the key themes that emerged from the development of the data source quality assessment framework was time. Time, in this context, refers to various temporal aspects such as when data was last updated, when it was edited, and how timely it is in relation to healthcare needs.

In data and knowledge representation, time refers to the temporal dimension that specifies when an event occurred, a value was recorded, or a state was valid [53]. Time can be represented as:

- Instant – a specific moment (e.g., "2025-04-23T10:00:00Z")
- Interval – a span between two time points (e.g., "April 2025")
- Duration – a measured length of time (e.g., "2 hours")

Time is essential for understanding trends, causality, validity, and synchronization across data sources [54]. During discussions with expert advisory groups involved in this framework design, the importance of time in data sources was a recurring theme that repeatedly highlighted its influence on data reliability and usefulness. The experts emphasised that the value of healthcare data significantly diminishes if it is outdated or not

promptly updated. In rapidly evolving healthcare environments, where decisions are often time-sensitive, the currency of data plays a pivotal role in clinical outcomes and treatment efficacy. Currency is a dimension of data quality that reflects how up-to-date the data is [55].

The experts also underscored the critical role of tracking the temporal dimension of data, noting how it can significantly improve data governance processes. When the time at which data is updated or edited is clearly recorded and tracked, it enhances the ability to audit data sources and ensure transparency. This clarity allows healthcare providers to better understand the reliability of the data they are using—knowing when the data was last modified can help ensure that decisions are being made based on the most up-to-date information available.

Moreover, tracking the temporal characteristics of data can improve overall data accuracy by allowing healthcare professionals to pinpoint when and how certain pieces of information were altered. This type of temporal traceability is essential for maintaining the integrity of healthcare records. For instance, if patient treatment records are updated at irregular intervals, it becomes challenging to track the progression of a disease or the effectiveness of a treatment plan. Temporal data tracking ensures continuity and consistency in patient care.

In the broader scope of data governance, the temporal dimension of data helps mitigate issues related to data mismanagement or inaccuracies. For example, regular updates to healthcare data sources can reduce the chances of errors or miscommunications that may arise from relying on outdated information. In this way, temporal data tracking also enhances accountability, as stakeholders can refer to the exact points in time when data was input or modified, ensuring that errors can be traced back and corrected efficiently.

Ultimately, the advisory group discussions revealed that time is not just a peripheral factor but a core component of healthcare data quality. It affects not only the relevance of the data but also its transparency, accuracy, and traceability. Incorporating a focus on temporal elements within the data source quality assessment framework helps to ensure that healthcare data remains a reliable and actionable asset in decision-making processes. This emphasis on time reinforces the need for timely, well-managed data systems that can adapt to the dynamic nature of healthcare.

In the next section temporal data modelling in healthcare is addressed in detail and different temporal data modelling approaches in healthcare are discussed.

In the next section, the FAIRVASC use case will be discussed, along with its application in this thesis. This will set the stage for its role in the research, illustrating how it supports the investigation of temporal data modelling in the healthcare domain.

2.3 FAIRVASC Use case

FAIRVASC is a research initiative by the European Vasculitis Society and the RITA European Reference Network, involving ten European partners focused on patient care for ANCA vasculitis [25, 26]. These partners include several national registries such as RKD[2] in Ireland, UKIVAS in the UK, and others from France, Czech Republic, Poland, Germany/Austria/Switzerland, and Sweden. The RKD registry, established in 2012, is dedicated to researching rare kidney diseases in Ireland and is a key contributor to this paper. The FAIRVASC ontology[3] standardises terminology across seven registries to manage ANCA vasculitis data effectively. Feedback from each registry helps refine these terms, which are incorporated into the ontology, covering aspects such as Patient, Diagnosis, and Clinical Outcomes.

Patients with ANCA vasculitis can experience relapse, which refers to the recurrence or worsening of disease symptoms after a period of improvements or remission. Understanding the specific medications used by patients during these critical periods is essential for designing more effective treatment plans and managing patient conditions more accurately. Without this temporal information, it is challenging to correlate medication use with patient outcomes, particularly in relation to relapse events [56].

This thesis focused on a subset of the RKD registry, containing patient details and medication records for 600 patients, stored in a relational database. The data included patient class which is linked to patient-related information like patient ID, gender, year of birth, and patient overview class. Patient overview class which is linked to relapse date and treatment class, and treatment class which contains details regarding the medications utilised for a patient as a treatment plan. The subset of ontology based on the dataset without including medication start date and stop date was considered as a base model. (Figure 6)

[2]https://www.tcd.ie/medicine/thkc/our-research/rita-ireland-vasculitis-riv-registry-and-biobank/
[3] https://ontologies.adaptcentre.ie/fairvasc/

The base model lacks the capability to present medication start and stop dates. This limitation is significant because each patient may use one or more medications during their disease progression, both before and after relapses. Modelling medication start and stop dates presents a significant challenge. Each medication has two associated dates: a start date and a stop date. The complexity increases when a patient uses the same medication at different intervals, resulting in multiple start and stop dates for a single medication. This situation complicates the modelling process, as it requires accurately capturing and representing these temporal relationships. To address this gap, the base model was extended to include temporal classes that represent medication start and stop dates and the proposed metrics framework was refined during the KG design and evaluation process.

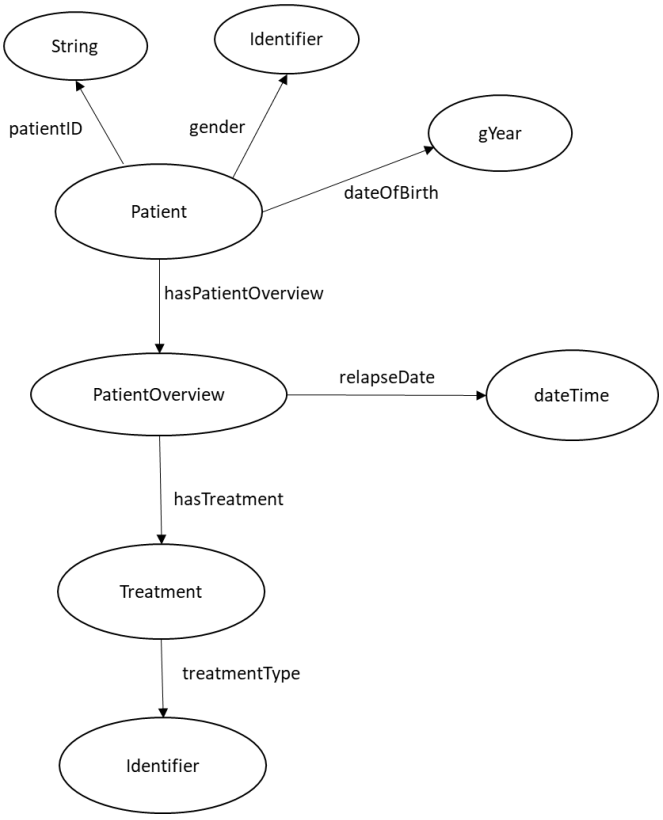


Figure 6- Base model - ontology without including medication time

In general, clinical registries present significant complexity due to variability in data collection practices, standards, and quality. Registries are often developed independently, tailored to specific clinical or research objectives, and therefore use different terminologies, coding systems, and data models to capture information about patients, treatments, and outcomes.

Concepts that seem universal, such as diagnoses, relapse events, or treatment responses, may be defined or interpreted differently across registries, making direct comparison and integration difficult. Furthermore, the depth and completeness of recorded data can vary widely: some registries capture detailed longitudinal histories including medication start and stop dates, dosages, and clinical measurements, while others may record only high-level information or omit temporal details altogether. Temporal complexity arises when patients receive multiple treatments over time, sometimes concurrently or intermittently, requiring sophisticated modelling of overlapping and sequential events. Data quality issues such as missing values, inconsistent date formats, and retrospective entry further complicate the use of registry data for structured analysis. Additionally, even within a single registry, updates to data collection forms and evolving clinical practices can introduce internal inconsistencies over time. These challenges necessitate careful harmonisation, standardisation, and flexible data modelling strategies to ensure that registry data can be reliably integrated, queried, and interpreted for research and clinical decision-making.

While the FAIRVASC initiative involves multiple national registries with varying data structures and standards, this thesis did not address the challenges associated with aligning data across these heterogeneous sources. Instead, the focus was on designing and evaluating data modelling approaches, particularly in relation to representing temporal aspects of clinical data. A single registry—the RKD registry in Ireland—was used as the primary data source to maintain consistency and reduce complexity. As such, issues such as semantic alignment, terminology differences, and integration of diverse data formats were outside the scope of this work. The framework developed was intended to explore how temporal information, such as medication start and stop dates, could be effectively modelled and evaluated within a controlled dataset.

2.4 Chapter Summary

This chapter reflects the essential role of time and temporal data in the context of healthcare data and data sources. It emphasises the importance of ensuring that healthcare data is not only accurate but also timely, highlighting the significant impact of outdated or delayed data on healthcare decisions and patient outcomes. The development of a data source quality assessment framework in this chapter underscores the need to consider both the content of the data and the reliability and timeliness of its sources.

Time emerges as a theme throughout the chapter, with experts stressing how critical it is to track when data is updated, edited, or used in decision-making processes. Delays in data can undermine the effectiveness of healthcare treatments and interventions, especially in situations where timely information is crucial for diagnosing conditions, tracking patient progress, or adjusting treatment plans.

The FAIRVASC use case illustrates this need by providing a real-world example of how temporal data modelling is applied to monitor patients with vasculitis. This use case not only demonstrates the challenges of managing complex, time-sensitive healthcare data but also shows how improved data models can help address these issues. The case highlights the importance of tracking patient information over time to ensure accurate and effective disease management.

In summary, Chapter 2 highlights the relationship between time and data quality in healthcare. The discussions on time emphasise the need for timely updates and accurate data to improve patient care and outcomes, while the practical application of the FAIRVASC use case demonstrates the value of temporal data modelling in addressing real healthcare challenges. This reflection on the importance of time sets the stage for deeper exploration of temporal data modelling in healthcare throughout the thesis.

Chapter 3 State of the art review

This chapter provides a comprehensive review of the current state of the art in modelling temporal data using KGs and in healthcare, addressing the second research objective, which aims to answer the corresponding research question.

The chapter explores various approaches to modelling temporal data in KGs and healthcare and discusses the state of art in evaluating different approaches. The goal is to offer a detailed overview of the existing approaches, identify research gaps, and highlight opportunities for further investigation in temporal data modelling and evaluation in KGs. This chapter lays the groundwork for understanding the advancements and challenges in the field, setting the stage for the research presented in subsequent chapters.

3.1 Temporal data modelling in healthcare

In this section, the process of conducting a literature review to identify temporal data modelling approaches in healthcare is discussed.

3.1.1 Motivation and scope

The amount and accessibility of data have grown significantly during the previous few years. This reality is mostly due to the growth of data in numerous formats from the most diverse areas (e.g. health care, banking, government or logistics) and sources (e.g. mobile devices, sensors, and social networks) that are constantly providing either structured, semi-structured, or unstructured data. As a response, data modelling and data analytics have emerged in a manner that allows us to process enormous amounts of data [18].

The information management community employs the term "data model" frequently, and it is used interchangeably with the term “database model” [57]. Data model has many various meanings. A data model, in its broadest meaning, is a set of conceptual tools used to represent real-world things and the connections between them [57, 58]. The phrase is also frequently used to describe a group of different data structure types [59]. Data models' main objective is to provide the structure, definition and format of data to assist in the design of databases.[60]

In healthcare, data models are used in designing databases for health information systems including electronic health records, personal health records, patients or disease registries,

and mobile health applications [61]. Healthcare data are generated over time by different stakeholders in healthcare including physicians, nurses, specialists, and other healthcare workers. The data originates from different heterogeneous sources such as hospitals, patient registries, clinics, and laboratories [19, 20].

Data that have time related attributes such as year, month, day, hour, minute and second are called temporal data [62]. Temporal data in healthcare are considered fundamental in prognosis, diagnosis and treatment process [19]. When determining a prognosis or a patient's diagnosis, temporal data such as encounter times, lab test times, medication consumption time, and follow up times must be recorded by the clinician. Modelling temporal healthcare data could benefit healthcare practitioners in healthcare decision making and support patient care in different ways including assessing the patient's clinical history, finding the possible causes of clinical events over time, exploring the temporal patterns in the diseases and identifying the disease progression and predicting the future trends and events based on the past and current clinical data [19, 63]. For example, in order to assess a patient's status, the clinician records the clinical history of a patient including laboratory and pathology results, symptoms, past diseases, therapies, medication used, patients' narratives and other information captured from patients, e.g., blood pressure and heart rate, which all contain temporal attributes [63]. The temporal changes in clinical parameters are critical to monitor patients status; for example monitoring the level of blood pressure and vital signs in patients in ICU [64].

There are mainly two types of temporal dimensions, valid time and transaction time. They refer respectively to the time that a fact is true and to the time that the data is created in the database. Valid time and transaction time can be represented as time-points and time-intervals. Time point refers to a point in time and time interval refers to a duration which includes start time and end time [62, 65, 66]. For example in some conditions the time point is used to refer to the time at which a lab result was measured specify a lab result measurement and in some conditions the time interval is used to define the duration of a therapy.

Previous researches in this area reviewed different aspects of temporal data modelling in healthcare; In 2006, a position paper regarding research directions and challenges in temporal representation and reasoning in medicine was published by Adlassnig et al [67]. In this paper relational databases were used for modelling temporal data in healthcare and their limitations and advantages were highlighted; however, there was no evidence of describing

the relational data modelling approach along with different time dimensions and healthcare applications.

In 2016, a review was performed, focusing on temporal data representation, normalisation, extraction and reasoning in clinical domain [61]. In this review, standards including HL7 v3, OpenEHR that are used for developing clinical information models, were described, and ontologies for modelling time and time expressions were explained; however, there was no systematic methodology for the review and the applications of temporal data modelling and different time dimensions were not addressed.

To our knowledge, there has been no review of available evidence to systematically investigate current approaches of temporal data modelling in healthcare along with the different time entities (valid time and transaction time), and applications of temporal data modelling in healthcare, describing the gaps and stating the future directions. Therefore, the aim of this research is to systematically describe current approaches and applications of modelling temporal data in healthcare, compare approaches in terms of application areas, time dimensions and time representations and describe the gaps. A list of recommendations for future research directions will be also provided.

The results of this study will inform data engineers, clinicians, health informatics researchers and others about the temporal data modelling approaches in healthcare when modelling temporal health data in designing healthcare databases and data repositories which support development of health information systems such as electronic health records, clinical decision support systems, health applications, and patient portals.

3.1.2 Method

The procedure that was undertaken in this study is as follows:

1. Formulating the research questions;
2. Selecting information sources (digital libraries) on which to perform search;
3. Defining search concepts and keywords;
4. Application of search terms on databases;
5. Considering inclusion and exclusion criteria for selection of studies;
6. Synthesis of data.

3.1.2.1 Formulating the research question

The literature review was conducted using the PEO (Patient/Population, Exposure, and Outcome) framework for developing the research question as follows, Population: Healthcare data, Exposure: Temporal, and Outcome: Model.

This study poses the following research questions: What are the existing approaches, and applications in modelling temporal data in healthcare?

3.1.2.2 Information sources

Searches were carried out on the following databases:

- IEEE Xplore Digital Library [68]
- PubMed

PEO framework	Concepts	Search terms
Population	Healthcare data	"clinical data" OR "Health care data" OR "Healthcare data" OR "health data" OR "medical data" OR "clinical information" OR "Health care information" OR "Healthcare information" OR "health information" OR "medical information"
Exposure	Temporal	Temporal OR "valid time" OR "transaction time"
Outcome	Model	model OR schema OR representation

3.1.2.3 Search strategy.

The search terms based on PEO framework, are set out in Table 3.

Table 3- Search terms for a systematic review on modelling temporal data in healthcare settings

The search query which was used for PubMed and IEEE Xplore, was:

("Clinical data" OR "Health care data" OR "Healthcare data" OR "health data" OR "medical data" OR "clinical information" OR "Health care information" OR "Healthcare information" OR "health information" OR "medical information") AND (Temporal OR "valid time" OR "transaction time") AND (model OR schema OR representation)

3.1.2.4 Inclusion and exclusion criteria

No limits were applied to articles for inclusion in terms of publication date or language. Articles were included if they discussed approaches of temporal data modelling in healthcare

settings. Studies were excluded if they referred to modelling temporal data in the non-health related data. Peer-reviewed literature were selected for review in this study.

3.1.2.5 Screening

All retrieved articles from the selected information sources were imported into Covidence. Duplicate references were removed. Two reviewers independently screened the titles and abstracts against the inclusion/exclusion criteria. Any disagreements on inclusion/exclusion were firstly resolved by discussion. Any disagreements not resolved by discussion were resolved by a third author.

3.1.2.6 Data extraction

Data extraction form built-in Covidence was used to record information from each included article. One researcher read the articles in full and manually extracted the data directly into the data extraction table. The reference, setting, and all mentions of data source quality items, were recorded. The data extraction form was then exported into Excel, version 15.0 by Microsoft Corporation, as tables, in order to perform analysis.

3.1.2.7 Selection of studies

Search queries returned 1,137 articles, of which 34 duplicate articles were removed. Screening of titles and abstracts led to the exclusion of 1056 articles and retention of 47 articles for full text review. After considering the inclusion and exclusion criteria, 20 articles were included in the literature review. (Figure 7)

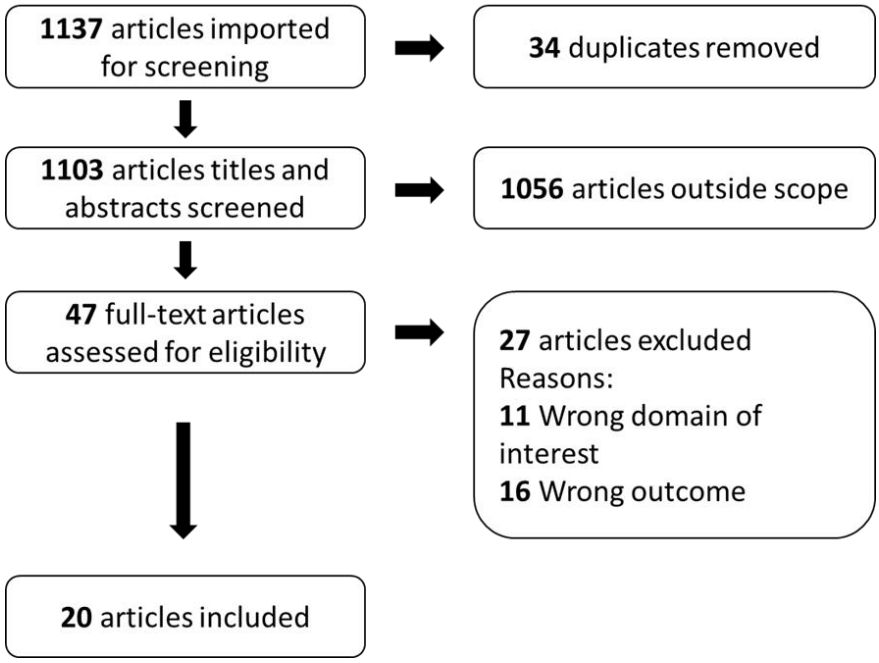


Figure 7-Flow chart of literature screening for review of temporal data modelling in healthcare

3.1.2.8 Data synthesis

The information was manually extracted from each included article. Articles were read in full by one researcher and the data extracted directly into the data extraction table. A narrative synthesis was performed to analyse the articles.

3.1.3 Results

The final 20 articles were included in this study for data analysis. (Appendix B) They published between 1990 and 2018. The findings have been analysed and categorised based on temporal data modelling approaches, applications and time dimensions and representations found in the literature.

Five different approaches for modelling temporal data were identified in the literature, including ontology, entity-relationship (ER) model, object-oriented model, relational model, and graph (knowledge graph) model. These approaches were classified based on three commonly recognized layers of data modelling: conceptual, logical, and physical, following the layered framework that includes external, conceptual, logical, and physical levels [64]. The classification is outlined as follows:

1. **Conceptual layer**: This layer provides an abstract, high-level view of the data and focuses on describing entities and relationships without technical implementation details. In this study, ontology and entity-relationship (ER) models were identified as conceptual modelling approaches.
2. **Logical layer**: This layer translates the conceptual model into structures that can be implemented in a database system. It includes more detail about data organization, types, constraints, and relationships. The object-oriented model, relational model, and graph data model (including knowledge graphs) fall under the logical layer. These approaches describe how data is structured and how relationships are managed in a system-independent manner.
3. **Physical layer**: While not the focus of this review, the physical layer refers to the actual storage of data on hardware, including indexing, file structures, and access paths. The studies reviewed typically did not address this level explicitly.

A description of each approach is outlined below:

1. **Ontology:** An ontology defines concepts and describes their relationships in a domain. The building blocks of an ontology includes: classes and concepts, properties, and attributes [69]. Classes like TemporalEntity, Instant, and Interval allow the representation of time-specific facts. In several disciplines, including the Semantic Web, e-commerce, and information retrieval, ontology provides a common and reusable body of knowledge about a particular topic [70, 71]. Based on the findings, this approach is used for querying, and analytics for the purpose of patient monitoring and follow up, medical record management, and clinical research.
2. **Entity relation (ER) data model:** An ER data model is most frequently used in database design. A relationship in an ER model represents a connection between two or more entities, whereas an entity in the model denotes a discrete object or concept. Attributes may well be found in both entities and relationships [72]. Temporal extensions to the ER model introduce constructs for handling time-varying attributes, historical relationships, and entity lifespans. This can include adding timestamped attributes, temporal entities, or temporal relationships. A common technique is to represent temporal aspects using start and end time attributes or to model historical versions of entities. In this study, the ER data model was proposed for modelling temporal data and it was used for querying for the purpose of medical record management.
3. **Object oriented data model:** The object oriented data model represents data and its relationship in an object which is a basic modelling structure [60]. Object-oriented temporal models encapsulate both the state and behavior of temporal objects. Time-dependent properties are managed through temporal classes, where objects maintain a history of states or versioned instances. Temporal inheritance and polymorphism are used to represent changes in data over time. For instance, a PatientVisit object could have temporal attributes capturing visit duration and associated health events. This approach is useful in managing evolving healthcare data in patient monitoring systems and clinical studies Based on the findings, this approach is used for querying, storage, and analytics for the purpose of patient monitoring and follow up, medical record management, and clinical research.
4. **Relational data model**: A relational data model represent data in the form of relations (tables). The table rows represent a real world entity. To interpret the significance of the values in each row, the table name and column names are useful [60]. Temporal relational models extend traditional relations by introducing valid-time, transaction-time, or bitemporal tables. These are implemented using additional timestamp columns

(valid_from, valid_to) to capture the temporal scope of each tuple. Techniques such as temporal normalization, tuple versioning, and temporal joins are applied for querying historical data. In the reviewed studies, this model supported querying and analytics of longitudinal patient records and medication history [55]. Based on the findings, this approach is mostly used for querying, analytics, and storage for the purpose of patient monitoring and follow up, medication management, and medical record management.

5. **Graph data model (knowledge graph):** A knowledge graph is a set of vertices (or nodes) and edges. The vertices represent concepts or entities and the edges shows the semantic relationships between concepts. Structured or unstructured information can be mined through knowledge graphs techniques and be integrated into knowledge and finally be represented in a graph or other similar output. In addition, knowledge graphs organise information in a manner that is easy-to-maintain, easy-to-understand and easy-to-use [73]. Knowledge graphs are mainly popular for integrating data from heterogeneous sources of data at large scale [74]. Based on the finding of this study, graph data models are mostly used for storage, analytics and querying temporal data in healthcare. This approach has been used for temporal data modelling in clinical research.

The articles were analysed based on modelling two time representations, time point and time interval and were categorised in four different categories which are outlined below.

1. Articles which focused on modelling time points
2. Articles which focused on modelling time intervals
3. Articles which focused on modelling both time points and time intervals
4. Articles which did not specify a time representation

Four application areas in healthcare were identified which used temporal data modelling approaches including patient monitoring and follow up, medication management, medical record management, and clinical research.

1. **Patient monitoring and follow up:** The focus of articles in this category was proposing a temporal data modelling approach for monitoring the patients' status and clinical follow up. Relational data model, object oriented data model and an ontology was proposed for querying the data and data analytics purposes.
2. **Medication management:** The focus of articles in this category was proposing a temporal data modelling approach for medication management. Relational data model, was proposed for data storage and analytics purposes.
3. **Medical record management:** The focus of articles in this category was proposing a temporal data modelling approach for medical record management. Object oriented

data model, ontology, entity relation (ER) data model, and relational data model were proposed for data storage, querying and analytics purposes.

4. **Clinical research:** The focus of articles in this category was proposing a temporal data modelling approach for clinical research. Object oriented data model, ontology and graph data model were proposed for data storage, querying and analytics purposes.

Table 4-Approaches of temporal data modelling in healthcare in relation to applications in healthcare in terms of functionality

Row	Applications in healthcare Approaches for temporal data modelling	Patient monitoring and follow up	Medication management	Medical record management	Clinical research
1	Proposing an ontology	Querying Analytics		Analytics	Analytics
2	Using Entity relationship data model			Querying	
3	Using Object oriented data model	Querying Analytics		Querying	Storage Analytics
4	Using Relational data model	Querying	Storage Analytics	Querying Storage	
5	Using Graph data model				Storage Analytics Querying

3.1.4 Discussion

In this study, a comprehensive review was performed on previous literature in the area of temporal data modelling in healthcare. A total of five approaches of modelling temporal data in healthcare were extracted from literature which were categorised in two different layers including conceptual layer and physical layer. Four applications in healthcare were identified. Two different time representations including time points and time interval were

extracted which were used to model temporal data in healthcare and the focus on both time representations were almost the same in the retrieved articles.

The approaches were categorised in two categories including conceptual layer and physical layer. In the conceptual layer, ontology is considered a more widely used approach to define concepts and relationships compared to ER data model. In the physical layer, relational data model and object oriented data model have been used in a wide range of healthcare applications compared to graph data model. The gaps and recommendations for future research are discussed below:

3.1.5 Identified limitations

1. One efficient and effective knowledge integration approach which has become prevalent in recent years is graph data models (knowledge graphs). Graph data models have attracted attention in several areas including natural language processing, question answering machines and recommendation systems. Methods to deal with and capture the variation and development of data over time in Knowledge graphs is of increasing interest and more research is needed to explore solutions for modelling temporal data in knowledge graphs in different healthcare applications including patient monitoring and follow up, medication management, medical record management and clinical research.
2. Temporal data modelling approaches have not been evaluated with different healthcare use cases and for data from heterogeneous sources. Existing approaches have produced different results in terms of performance in querying, storage efficiency and usability which still needs to be further evaluated and compared.

3.2 Systematic review

To identify approaches, applications of temporal data modelling in KGs, a systematic review was conducted. The detail of this review is discussed below:

3.2.1 Methods

The review's notable strength lies in its rigorous methodology for conducting a systematic review to identify temporal data modelling approaches in KGs. This systematic review is based on the guidelines and procedures for conducting systematic reviews within the software engineering domain proposed by Kitchenham [75]. This guideline has significantly contributed to the advancement of systematic methods in the realm of computer science literature reviews.

The majority of existing reviews in computer science are not performed systematically and do not contain the steps mentioned in this study. Most of the reviews in computer science are called "survey", and do not follow a rigorous method and this may affect the outputs. Preparing a protocol for the systematic review and following the systematic review guidelines assist in finding comprehensive results. It also helps to make the work replicable for other research. Researchers can use the terms and strategy and run the search in the future in order to update the findings to reflect changes in the evidence base. This creates efficiencies for the computer science research community. This way of performing systematic reviews also encourages the standardisations of terms used in the community to enhance searching the literature.

3.2.1.1 Search Strategy

The research questions is, what are the existing approaches in modelling temporal data in RDF based knowledge graphs.

Searches was carried out on the following databases: ACM Digital Library [76], IEEE Xplore Digital Library[68], and ScienceDirect [77]. The bibliographies of the included full-text articles was searched for relevant articles. Searching of forward citations was also conducted to identify other potential material for inclusion. The search terms used, are set out in Table 5.

Table 5-Search terms for a systematic review on modelling temporal data in knowledge graphs

Concepts	Search terms
Knowledge graph	"Knowledge graph" OR RDF OR "resource description framework"
Temporal	Temporal OR dynamic OR evolution OR time
Applied model	*present OR annotate OR model OR schema OR standard OR framework OR structure OR application OR applied

The search query to be used for ACM digital library, as an example of the search strategy, is:

*"("Knowledge graph" OR rdf OR "resource description framework") AND (Temporal OR dynamic OR evolution OR time) AND (*present OR annotate OR model OR schema OR standard OR framework OR structure OR application OR applied*

The search query is designed to capture a wide range of relevant literature on temporal knowledge graphs, and while it may include some broad terms, this strategy enhances the likelihood of uncovering a diverse array of studies. By using keywords like "model," "schema," and "framework," the query ensures that it encompasses various approaches to temporal knowledge graph applications, allowing for a comprehensive view of existing methodologies. The use of the truncation operator (*present) is particularly beneficial for capturing different word endings, ensuring the search remains inclusive of various terminology. Although the query may return some results outside the immediate focus, it fosters a broader exploration of related topics and promotes discovery of potentially valuable but less obvious sources.

Two reviewers independently screened titles and abstracts based on the inclusion and exclusion criteria. Any disagreements not resolved by discussion was resolved by a third author. Forward citation and hand-searching of bibliographies of included studies was performed and any relevant studies identified were included. Articles selected for full text screening were screened by one reviewer. Publication date and language limits were not applied.

3.2.1.2 Study Selection criteria

Original research articles proposing approaches in modelling temporal data in KGs or discuss applications of temporal data modelling in KGs or address challenges of temporal

data modelling in KGs were reviewed. Systematic reviews, surveys, literature reviews were excluded but the reference lists were screened for the potential studies fitting the inclusion criteria. Studies were excluded if they referred to knowledge graphs based on frameworks other than Resource Description Framework (RDF).

3.2.1.3 Data extraction

A data extraction table was created using the data items extracted from the studies that were included, and this table was then exported to Microsoft Excel [50]. To ensure appropriateness of the included data extraction fields against the data provided in articles and mutual understanding of the fields, a data extraction table was first piloted on three articles by two researchers and then updated and more columns were added. The final table included article title, RDF/S extension, the semantics of modelling, time dimensions used for modelling, applications and use cases used in the articles.

3.2.1.4 Quality Assessment

All studies were included in the synthesis since any finding related to the research objective improves the synthesis of the final evidence.

3.2.1.5 Data Synthesis

A narrative synthesis of the papers was conducted to summarise the findings. The articles were categorised based on modelling approaches in different categories including design approach, semantics, and time dimensions. Each category was explained in detail and the approaches within each category were introduced. In order to assist with information visualisation, the synthesised data was presented in a range of formats, including tables, graphs, and other visual representations.

3.2.2 Results

The initial search retrieved 8479 papers, of which 1376 duplicate papers were removed. Screening of titles and abstracts led to the exclusion of 6965 papers and retention of 138 papers for full text review. After considering the inclusion and exclusion criteria, 30 papers were included in the literature review. 7 papers were also identified for inclusion based on hand searching and forward citation. Finally, 37 papers were included in the review. (Figure 8) The process was performed using the Covidence systematic management software. The main reasons for excluding studies were: research did not focus on RDF; and not proposing a temporal data modelling approach in KG [49].

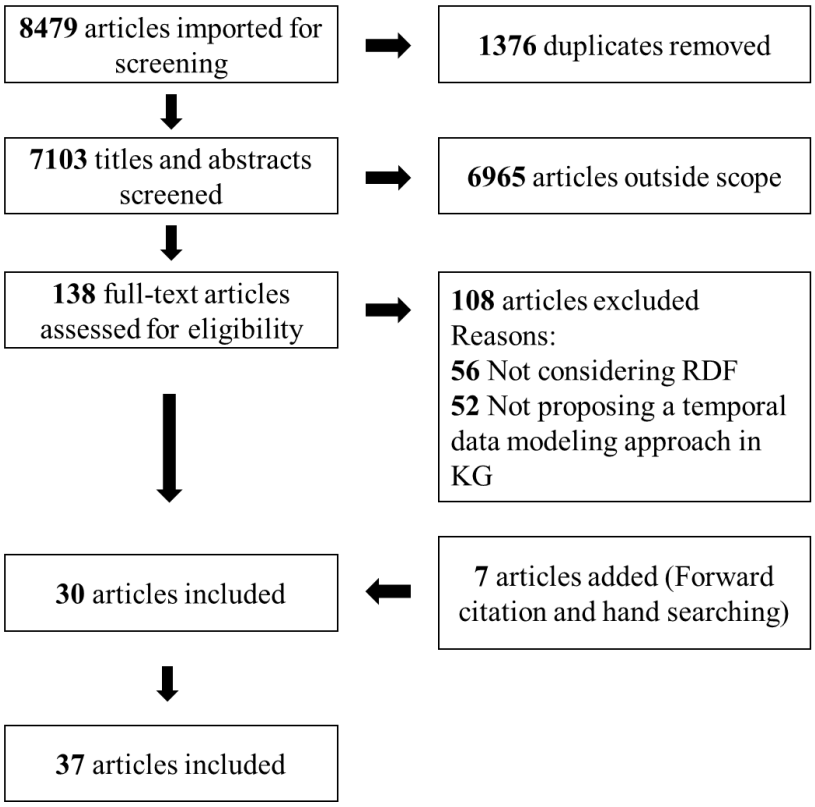


Figure 8-Flow chart of literature screening for systematic review of temporal data modelling in KGs

3.2.2.1 Approaches

The extracted approaches were categorised based on semantics (RDF/S extension, OWL, Logic), time dimensions, scope of applications, and use cases. (Table 1) From 37 papers, 33 unique approaches were identified. The approaches cover a wide range of use cases/datasets, from flight information to biomedical knowledge repositories to web services. Some of the approaches are focused purely on temporal modelling, while others are more geared towards spatiotemporal modelling or meta-knowledge modelling. Some of the approaches also consider transaction time, in addition to valid time.

Table 6- Comparison between approaches of temporal data modelling in knowledge graphs

No	Approach name	Semantics			Time Dimension		Scope of application	Use case/datasets
		RDF/S Extension	OWL	Logic	Valid Time	Transaction time		
1	stRDFS[78]	✓			✓		Spatiotemporal modelling	Flight information-spatiotemporal data
2	RDFMT[79]	✓			✓		Temporal modelling	None
3	RDFM[80]	✓			✓		Metadata modelling	Biomedical Knowledge Repository (BKR) dataset a Gov-track dataset a Synthetic dataset
4	UTKGs[81]			✓	✓		Temporal modelling	footballdb.com and wikidata.org
5	The lightweight biography ontology[82]		✓		✓		Temporal modelling	Biography information
6	Dimensional Ontology (DO)[83]		✓		✓		Temporal modelling	E-Government data
7	A logic-based approach for			✓	✓		Temporal modelling	Food data set

No	Approach name	Semantics			Time Dimension		Scope of application	Use case/datasets
		RDF/S Extension	OWL	Logic	Valid Time	Transaction time		
	representing validity time in RDF(S) and OWL 2.[84]							
8	CNTRO[85]		✓		✓		Temporal modelling	clinical narratives
9	stRDF[78]	✓		✓	✓		Temporal modelling	Semantic Sensor Web
10	NG-tempRDF[86]	✓			✓		Temporal modelling	software source code information A data set describing Swiss parliamentarians
11	Multi-temp RDF [87]	✓	✓		✓	✓	Metadata modelling	None
12	Labeled RDF[88]			✓	✓		Metadata modelling	None
13	temporal extension of the Hayes and ter Horst entailment rules for			✓	✓		Temporal modelling	None

No	Approach name	Semantics			Time Dimension		Scope of application	Use case/datasets
		RDF/S Extension	OWL	Logic	Valid Time	Transaction time		
	RDFS/OWL [89]							
14	UncertainTemRDF[90]			✓	✓		Temporal modelling	T-YAGO dataset
15	Onto-temporal RDF[91]			✓	✓		Temporal modelling	None
16	RDFt[92]			✓	✓		Temporal modelling	two datasets of NBA players and YAGO
17	Singleton property [93]	✓			✓		Metadata modelling	The Biomedical Knowledge Repository (BKR) and YAGO
18	tRDF[94]			✓	✓		Temporal modelling	the GovTrack dataset consists over 20.5M triples of public record
19	An upper-level ontology for modelling thematic, temporal and spatial dimensions of data.[95]		✓		✓		Spatioetemporal modelling	None

No	Approach name	Semantics			Time Dimension		Scope of application	Use case/datasets
		RDF/S Extension	OWL	Logic	Valid Time	Transaction time		
20	Annotated RDF [96]			✓	✓		Metadata modelling	Synthetic data
21	Temporal RDF graph [97]	✓			✓		Temporal modelling	Description of Web services.
22	TA-RDF[98]	✓			✓		Temporal modelling	Sensor data
23	TKB[99]	✓			✓		Temporal modelling	knowledge base of a historian that records information about countries and their political governance
24	TempCRM[100]			✓	✓		Temporal modelling	smart house
25	MRDF [101]	✓			✓		Metadata modelling	None
26	RDF-Star[102]	✓			✓		Metadata modelling	None
27	RDF+[103]	✓			✓		Metadata modelling	Web pages of Computer Science departments
28	N-ary relations [104]		✓		✓		All Meta knowledge modelling	None

No	Approach name	Semantics			Time Dimension		Scope of application	Use case/datasets
		RDF/S Extension	OWL	Logic	Valid Time	Transaction time		
29	4d fluent [105]		✓		✓		Metadata modelling	None
30	SOWL [106]		✓		✓		Spatiotemporal modelling	spatial and spatio-temporal information
31	towl [107]		✓		✓		Temporal modelling	leveraged buyout (LBO)
32	named graph[108]	✓			✓		Metadata modelling	None
33	Standard Reification[109]	✓			✓		Metadata modelling	None

3.2.2.2 Categories

A summary of each category is discussed below:

Semantics:

- *RDF/S Extension:* This category extends the RDFS (RDF Schema) language with time-related properties and classes, allowing temporal information to be expressed in a more natural way within the RDF data model.
- *OWL:* OWL is a powerful and expressive Semantic Web language for representing ontologies and knowledge graphs. It can also be used to represent temporal information in a knowledge graph. This is accomplished by creating temporal classes and properties to represent time intervals, durations, and other temporal entities. 9 approaches were found in this category.
- *Logic:* Logic-based approaches can also be used to add temporal information to a knowledge graph. Usually, this entails defining temporal operators that can be used to

reason about temporal relationships between entities. Allen's interval algebra, for example, provides a set of temporal operators that can be used to reason about temporal relationships between intervals.

Time Dimension: Time in knowledge graphs can be represented in two ways: valid time and transaction time which are discussed below.

- *Valid time:* Valid time refers to the period during which a fact or statement is true. For example, in a knowledge graph about historical events, a fact about a specific event may have a valid time of the period during which the event took place.
- *Transaction time:* Transaction time, on the other hand, refers to the time at which a fact or statement is recorded or becomes known. For example, in a knowledge graph about medical records, a fact about a patient's health may have a transaction time of the moment it was recorded in the system.

Scope of application: In this category, approaches were categorised based on different scopes of applications including specifically for temporal modelling, specifically for spatiotemporal data modelling, and meta-knowledge representation in general. (Figure 9)

- *Temporal Modelling:* Temporal modelling in knowledge graphs involves representing and reasoning about time-varying data. This can include events, time intervals, and other time-dependent phenomena. Temporal modelling can be used in a wide range of applications, from predicting future events to identifying patterns and trends in the data.
- *Spatiotemporal Data Modelling:* Spatiotemporal data modelling is a specialised application of temporal data modelling that involves representing and reasoning about both temporal and spatial relationships between entities. This can be particularly useful in applications such as geographic information systems, where the location and timing of events are both important.
- *Metadata Modelling:* Metadata modelling is a broad application of modelling data in KGs and involves representing knowledge about the structure and semantics of a KG itself. This can include information about the entities and relationships in the graph like time, as well as information about the sources and provenance of the data.

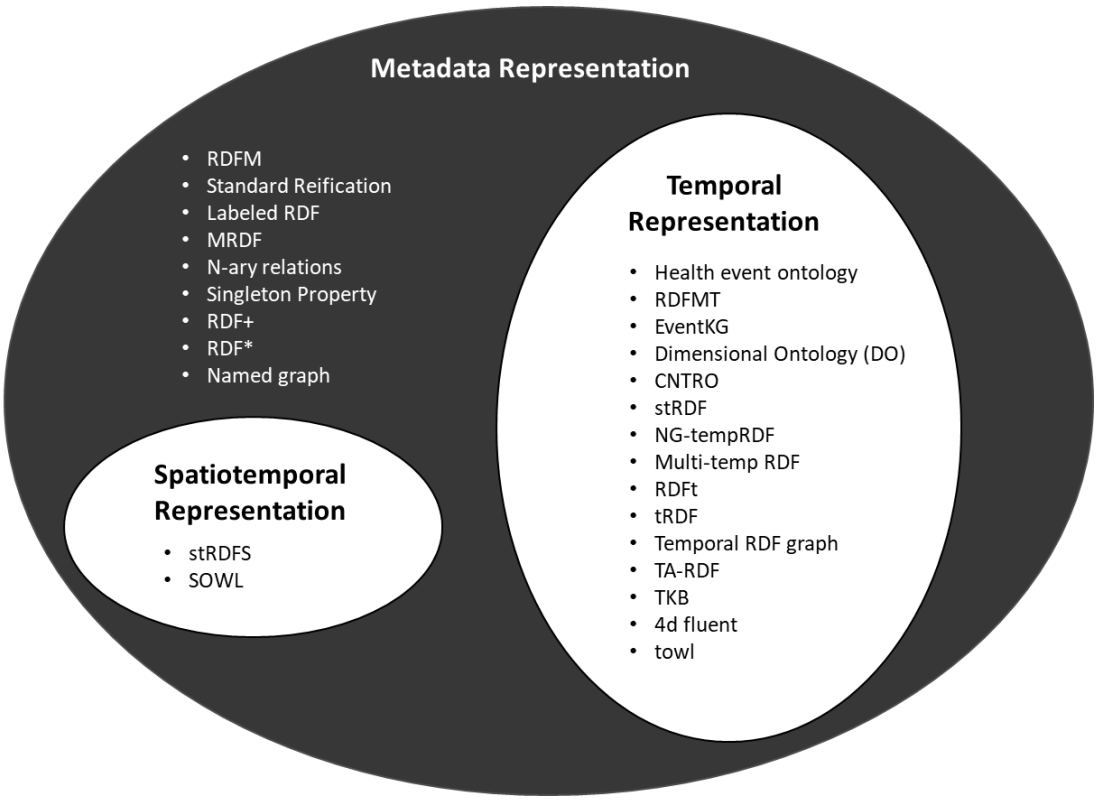


Figure 9- Temporal data modelling approaches in KGs in terms of scope of application

Use case: In the context of modelling temporal data in KGs, a use case refers to a specific application or scenario in which temporal data modelling is used to achieve a particular goal or solve a particular problem.

Based on the findings of this study, some use cases have been used for modelling temporal data in KGs. The use cases/datasets were categorised based on the similar characteristics and are outlined below:

1. *Spatial/Spatiotemporal Data:* This category includes datasets that model relationships between location and time. The Flight information-spatiotemporal dataset tracks the arrival and departure times of flights at different airports, while the Semantic Sensor Web dataset provides information about sensors that measure environmental conditions at different locations over time. The Spatial and spatio-temporal information dataset may contain a variety of spatial and spatio-temporal data, including geographic coordinates and time stamps.

2. *Healthcare:* This category includes datasets related to medical and healthcare information. The Biomedical Knowledge Repository (BKR) dataset contains information about diseases, treatments, and patient histories, while the Clinical Narratives dataset includes free-text medical records that can be analyzed for patterns and trends.
3. *Government Data:* This category includes datasets related to government activities and policy. The Gov-track dataset provides information about the activities of the U.S. Congress, while the E-Government dataset includes information about government services provided online. The dataset describing Swiss parliamentarians provides information about Swiss political representatives.
4. *Sports Data:* This category includes datasets related to sports and athletics. The Footballdb.com dataset provides information about football (soccer) players and teams, while the NBA player dataset tracks performance statistics for basketball players.
5. *Web Data:* This category includes datasets related to the web and online content. The web pages of Computer Science departments dataset contains information about academic departments at different universities, while the description of web services dataset models relationships between different online services.
6. *Historical Data:* This category includes datasets related to historical events and information. The Knowledge base of a historian dataset records information about countries and their political governance over time, while the Biography information dataset may contain biographical information about individuals throughout history.
7. *Other:* This category includes datasets that do not fit neatly into any of the other categories. The synthetic dataset is an artificially generated dataset used for experimentation, while the Food dataset contains information about different types of food. The software source code information dataset provides information about computer programming code, while the T-YAGO and YAGO datasets provide general knowledge information. The Smart house dataset contains information about home automation and the Leveraged buyout (LBO) dataset provides information about financial transactions.

3.2.2.3 Approaches

Below is a brief description of several approaches identified during the literature review, which were later applied in the experimental section.

Approach 1: Named Graph

The Named Graph approach in RDF allows for grouping triples into distinct subgraphs, each identified by a unique name or URI. This method is particularly useful for scenarios where data needs to be managed or queried in isolated contexts, such as different temporal snapshots, versions, or sources of data.[108]

Approach 2: Standard Reification

Standard Reification involves representing statements about statements, allowing the inclusion of metadata or contextual information about a specific triple in the graph. This method is foundational for handling temporal data in knowledge graphs, as it enables the capture of time-related attributes directly linked to individual triples. The detailed steps involve defining the reified triples, associating them with temporal metadata, and integrating them into the overall.[109]

Approach 3: RDF-Star

RDF-Star (also known as RDF*) is an extension of the RDF data model that allows embedding triples within other triples. This approach simplifies the representation of metadata (such as temporal information) associated with RDF triples without requiring complex reification structures. It is particularly useful for cases where additional context or attributes need to be directly linked to a specific triple.[102]

Approach 4: Temporal RDF Graph

The temporal RDF graph introduces a statement node to encapsulate time-dependent facts, such as a patient's treatment. This node is linked to a Temporal# node, which provides the time context of the statement. The Temporal# node is further connected to an Interval# node that defines the start and end dates of the time period using dateTime properties. This structure allows precise representation of when a specific fact or event is valid, making it possible to handle time-bounded information in RDF data, such as medical treatments or diagnoses over specific time intervals.[110]

Approach 5: RDFMT

The RDFMT approach introduces a special node called statement, which is assigned to the graph to represent the relationship between subject, predicate, and object (i.e., a triple). In this approach, temporal data such as startDate and stopDate are linked to each node of the

triple, allowing for the precise representation of when a particular fact (e.g., a treatment for a patient) is valid over time.[111]

Approach 6: RDFt

In this model, a statement node is introduced to represent a specific triple (subject, predicate, object), which allows for the inclusion of metadata attached to the predicate. [112]

Approach 7: Singleton Property

The singleton property approach in RDF is a method used to associate specific metadata, such as time, provenance, or certainty, with individual instances of RDF triples. It works by creating a unique property for each instance of a relationship (triple), allowing metadata to be linked to that specific relationship without altering the standard RDF triple structure. [113]

3.2.3 Discussion

This systematic review synthesises the evidence from 37 primary studies. Three semantic categories were identified based on the approaches. RDF/S extension, OWL, and logic-based approaches are all methods for adding time to a KG on the Semantic Web, although their fundamental ideas and capabilities differ. RDF/S extension is a relatively simple approach that can be easily integrated with existing RDF/S-based KGs, while OWL provides a more powerful and expressive approach to representing temporal information in a KG. In addition, logic-based approaches provide a formal and precise way to reason about temporal relationships in a knowledge graph. In summary, RDF/S extension is a simple and easy-to-use approach, while OWL provides more powerful and expressive capabilities, and logic-based approaches provide formal and precise reasoning capabilities. The choice of approach depends on the complexity of the temporal relationships that need to be represented and the available modelling and reasoning resources.

In terms of time dimension, valid time modelling tends to be more commonly used in applications where it's important to track changes over time and maintain a historical record of the state of the data. For example, in domains such as finance, healthcare, and legal records, valid time modelling is often used to ensure accurate and auditable records of events and transactions over time. Transaction time modelling, on the other hand, is typically used when the focus is on the state of the data at a specific point in time, and when there is less

need for a complete historical record. In applications such as monitoring and logging systems, transaction time modelling may be more appropriate.

In KGs, metadata representation of data refers to the representation of additional information about the data stored in the graph. This additional information can include various aspects such as time, trust, provenance and more. Meta data representation can include temporal data modelling and spatiotemporal data modelling, but it is not necessarily limited to those aspects. When modelling temporal data, it's important to capture the timing of events and entities in the graph. This can be represented using time intervals, timestamps, or a combination of both. In KGs, temporal data modelling can be used to support various tasks such as historical analysis, event sequencing, trend analysis, and forecasting. Spatiotemporal data modelling involves representing information about both the location and time of events or entities. This can be represented using spatial and temporal coordinates, geospatial data, or other location-based information. In KGs, spatiotemporal data modelling can be used to support various tasks such as route planning, disaster response, environmental monitoring, and epidemiology.

Based on the provided list of use cases and datasets, there is a wide range of potential applications for modelling temporal data in KGs and the best approach depends on the specific use case and requirements, and further exploration of these areas could lead to new insights and discoveries.

Temporal data modelling in KGs is essential for understanding and analysing time-dependent phenomena, making informed decisions, and predicting future events. By incorporating temporal data into a KG, we can create a more complete and accurate representation of the data and enable more powerful and sophisticated analysis and decision-making.

The approaches of temporal data modelling in KGs and their categories which discussed in this systematic review can be used to assist in making decisions on the design of KG based databases, repositories, and registries which contribute in developing Information systems and enhance data analytics.

Temporal data modelling in KGs has numerous potential applications, including finance, healthcare, and transportation. Future research could dive deep into these applications and create use cases that demonstrate the benefits of temporal data modelling in knowledge graphs based on RDF. Furthermore, existing approaches of temporal data modelling in KGs

produce different results in terms of performance in querying, storage efficiency, expressiveness, complexity, and usability which still needs to be further evaluated and compared with each other.

In the next section, existing literature regarding the evaluation of temporal data modelling approaches sin KGs are addressed.

3.3 Time Ontologies

3.3.1 OWL-Time Ontology

OWL-Time Ontology is a standardized vocabulary for describing temporal concepts in a structured and machine-readable way, using the Web Ontology Language (OWL) [114].

OWL-Time is used to model time-related information in data, such as:

- Dates and times
- Time intervals
- Durations
- Temporal relationships (e.g., one event happens before or after another)

It enables consistent and semantic handling of time-based information across systems, especially in Semantic Web, Linked Data, and ontology-driven applications.

Key Concepts include:

- time:Instant: A specific moment in time (e.g., 2025-04-23T10:00:00Z)
- time:Interval: A duration between two points in time
- time:hasBeginning: Start of an interval (linked to an Instant)
- time:hasEnd: End of an interval (linked to an Instant)
- time:duration:The length of a time interval
- time:before / time:after: Temporal ordering between events or intervals

This ontology aims to provide a standardized vocabulary for describing temporal concepts like instants, intervals, durations, and their relationships. In contrast, our research focuses on examining various design patterns for integrating temporal dimensions into knowledge graphs, with an emphasis on how these patterns manage time-based data and enable temporal reasoning.

3.3.2 W3C Data Cube Vocabulary

The W3C Data Cube Vocabulary is a standard for publishing multi-dimensional statistical data using RDF. It is particularly well-suited for structured datasets where time is a regular

and repeating component, such as census data, health statistics, or economic indicators [115].

In this model, time is represented as a dimension of an observation, using the qb:DimensionProperty. Each observation can include a time-related value—such as a specific year, month, or date—which describes when the data point was measured or recorded. For instance, a population statistic might be linked to the year 2025 using a property like ex:timePeriod.

Although the vocabulary itself does not enforce detailed temporal semantics, it can be integrated with other ontologies such as OWL-Time to provide richer representations of temporal intervals, instants, and relationships. This allows data publishers to distinguish between various levels of temporal granularity (e.g., annual, monthly) and even define custom temporal hierarchies if needed.

The Data Cube approach emphasizes the structural organization of data, making it suitable for analysis, visualization, and aggregation over time. Time is just one of many dimensions—others might include geographic location or demographic category.

3.3.3 W3C Semantic Sensor Network (SSN) Ontology

The W3C Semantic Sensor Network (SSN) Ontology, along with its lightweight core SOSA (Sensor, Observation, Sample, and Actuator), is designed to model sensor data, observations, and related metadata in domains like IoT, environmental monitoring, and smart cities [116].

In SSN/SOSA, time is a core attribute of an observation. There are two main temporal properties:

- sosa:resultTime – indicates when the result of an observation was produced or made available.
- sosa:phenomenonTime – refers to when the observed phenomenon actually occurred (if known).

This distinction allows for modeling real-world scenarios where there might be a delay between an event and when it is recorded (e.g., satellite imagery captured on one day but analyzed later). Observations in SSN are typically event-based or streamed, with irregular or real-time time intervals.

SSN is compatible with OWL-Time for advanced temporal modeling, such as defining intervals, durations, or temporal relationships between observations. This is especially

important when reasoning over time-series data or aligning observations from multiple sources.
Overall, SSN is designed for flexible, fine-grained temporal modeling in dynamic systems, where time is not just a data label but a fundamental aspect of observation.

These ontologies aim to provide a standardized vocabulary for describing temporal concepts like instants, intervals, durations, and their relationships. In contrast, our research focuses on examining various design patterns for integrating temporal dimensions into knowledge graphs, with an emphasis on how these patterns manage time-based data and enable temporal reasoning.

3.4 FAIR Data Principles

The FAIR data principles—Findable, Accessible, Interoperable, and Reusable—serve as guidelines to enhance the management and stewardship of digital assets. They emphasize machine-actionability to support automated data discovery and processing, which is crucial for effective temporal data modelling in knowledge graphs [117].

- Findable: Assigning globally unique and persistent identifiers to data and metadata ensures they can be reliably located and referenced.
- Accessible: Data should be retrievable by their identifier using standardized communication protocols, ensuring consistent access.
- Interoperable: Utilizing shared vocabularies and ontologies, such as OWL-Time and SSN, facilitates data integration and interoperability across systems.
- Reusable: Providing rich metadata and clear usage licenses enhances the reusability of data, allowing it to be applied in various contexts.

Implementing these principles in temporal knowledge graphs ensures that time-stamped data is not only structured and semantically rich but also discoverable and usable across diverse platforms and applications. While the FAIR principles provide a valuable framework for data management, this research primarily focuses on exploring design patterns for integrating temporal dimensions into knowledge graphs. The discussion of FAIR principles here serves to contextualize the broader landscape of data interoperability and reusability.

3.5 Ontology Design Patterns (ODPs)

Ontology Design Patterns (ODPs) are reusable solutions to common modeling challenges in ontology engineering. They encapsulate best practices for representing recurring structures

and relationships within ontologies, facilitating consistency and interoperability across different domains [118].

Key Characteristics of ODPs:

- Reusability: ODPs provide standardized templates that can be applied across various ontologies, promoting uniformity.
- Modularity: They support the construction of complex ontologies through the combination of smaller, well-defined patterns.
- Domain Independence: While some ODPs are domain-specific, many are designed to be applicable across multiple domains.

This research primarily concentrates on exploring and evaluating Knowledge Graph Design Patterns (KGDPs) for integrating temporal dimensions into knowledge graphs. While ODPs offer foundational modeling techniques, the emphasis here is on practical design patterns that facilitate the representation and reasoning of time-based data within knowledge graphs.

3.6 R2RML (RDB to RDF Mapping Language)

As the adoption of the Semantic Web grows, integrating existing relational data into RDF (Resource Description Framework) format becomes increasingly important. RDF mapping languages provide a formal mechanism to transform structured data—particularly from relational databases—into RDF graphs, enabling interoperability, reasoning, and semantic querying via SPARQL.[119]

One of the most widely accepted standards for this purpose is R2RML (RDB to RDF Mapping Language). R2RML was developed by the W3C RDB2RDF Working Group and became a recommendation in 2012. It defines a declarative, flexible language for expressing customized mappings from relational database schemas to RDF datasets.

R2RML mappings are written in RDF themselves, typically using Turtle syntax, and consist of Triples Maps, which describe how database rows are transformed into RDF triples. Each Triples Map includes:

- Logical Table: Specifies the source table or SQL query.
- Subject Map: Defines how to generate the RDF subject from one or more columns.

- Predicate-Object Maps: Describe how columns map to predicates and objects in RDF triples.

This structured and modular approach allows for complex mappings, including joins across multiple tables, URI generation, and the creation of RDF links between datasets.

R2RML has been widely adopted in research and industry. Despite its flexibility, one of the challenges of R2RML is its verbosity and steep learning curve, especially for users without prior experience in Semantic Web technologies.

As data integration and semantic interoperability remain critical in domains such as healthcare, finance, and e-government, R2RML continues to play a central role in enabling the transition from legacy relational systems to RDF-based knowledge graph.

3.7 Evaluation approaches in KGs

This section presents a comprehensive overview of evaluation and benchmarking approaches for RDF-based knowledge graphs (KGs), with particular attention to modelling strategies and their performance implications. The literature on RDF systems provides robust evaluation methods for SPARQL query performance and triple store capabilities, yet there remains a gap in evaluating modelling techniques, especially those dealing with temporal data. The aim here is to bridge this gap by synthesizing existing benchmarks, identifying their limitations, and highlighting opportunities for more holistic evaluation frameworks.

Overlaps and Distinctions in RDF Data Modelling, Knowledge Modelling, and RDF Data Quality

While RDF data modelling, knowledge modelling, and RDF data quality are closely related, they address distinct aspects of knowledge graphs and their application. RDF data modelling primarily focuses on the structure and representation of data in RDF format, which involves converting data into triples (subject-predicate-object) that can be easily queried using SPARQL. This type of modelling ensures data is stored in a way that is machine-readable and compliant with the RDF specification.[120]

In contrast, knowledge modelling is a broader concept that deals with how domain-specific knowledge is represented and structured using formal ontologies. This includes not only the data but also the relationships and rules that govern how data is interconnected. Knowledge modelling typically uses RDF but can also employ more advanced technologies like OWL (Web Ontology Language) to capture more complex relationships, enabling reasoning over the data.[121]

RDF data quality, on the other hand, evaluates the inherent quality of the data within an RDF graph, such as its accuracy, completeness, consistency, and timeliness. It measures how well the RDF data reflects the real-world entities and events it represents, and whether the data is reliable enough to support accurate analysis and decision-making. This is particularly important in applications like clinical data modelling or any use case requiring high levels of data integrity.[122]

Although these areas often overlap—especially when it comes to semantic representation in RDF—each has its distinct focus. RDF data modelling emphasizes the structural integrity of data, knowledge modelling ensures that the data accurately represents real-world concepts and relationships, and RDF data quality guarantees that the data is trustworthy and fit for use.

3.7.1 Benchmarking RDF and SPARQL Systems

A substantial body of research has established benchmarks to evaluate the performance of RDF stores and SPARQL engines. These benchmarks typically focus on query execution time, scalability, and storage efficiency, and are designed to assess system-level behaviors under controlled or diverse workloads.

Key benchmarks include:

- LUBM (Lehigh University Benchmark): Focuses on reasoning performance in ontology-based university data. It uses a synthetic generator to produce scalable datasets and a standard set of SPARQL queries [123].
- BSBM (Berlin SPARQL Benchmark): Simulates an e-commerce scenario and emphasizes business intelligence queries, covering both read and write operations [124].

- WatDiv (Waterloo SPARQL Diversity Test Suite): Known for generating diverse query shapes (linear, star, snowflake, complex), WatDiv tests system performance across a range of query types and data distributions [125].
- SP2Bench: Based on the DBLP computer science bibliography, this benchmark offers realistic document and citation data and emphasizes complex queries and joins [126].
- FedBench: Targets federated query processing, testing how well systems integrate and execute queries over distributed RDF datasets [127].

3.7.2 Evaluation of RDF Modelling Approaches

Evaluation of RDF data modelling techniques is less systematically developed than query or system benchmarks. This area concerns how different methods of representing knowledge—particularly metadata and temporal information—impact RDF system behavior, maintainability, and semantic clarity.

Several studies have addressed this challenge:

- Orlandi et al. introduced the REF benchmark, comparing RDF reification, singleton properties, and RDF-star. The benchmark evaluates the trade-offs in expressiveness, query complexity, and storage overhead when adding metadata to triples [15].
- Abuoda et al. developed StarBench to benchmark RDF-star and SPARQL-star features across multiple triple stores, emphasizing runtime performance and metadata representation [128].
- Fu et al. focused on provenance modelling techniques, including nano-publications and N-ary relations, using graph size and query processing time as evaluation metrics.[129]
- Hernandez et al. evaluated RDF modelling approaches using Wikidata, analyzing representation choices and performance trade-offs in provenance encoding.[109]
- Frey et al. conducted broader comparisons using real-world datasets like DBpedia and Wikidata, measuring statement counts, load time, and query efficiency across different metadata representation methods.[130]

While these studies provide useful insights, most rely on limited metrics (e.g., query response time, triple count), and few employ systematic or standardized methodologies for evaluating the full impact of modelling choices.

3.7.3 Evaluation of Temporal RDF Models

Temporal RDF modelling has been an active research area, especially in applications such as clinical data, longitudinal research, and event histories. Modelling temporal information in RDF often requires extending the base model with ontologies (e.g., OWL-Time), patterns (e.g., 4D Fluents), or metadata mechanisms (e.g., named graphs, RDF*).

Significant studies include:

- Wang and Tansel reviewed temporal RDF and OWL models, comparing the complexity and expressiveness of 4D Fluents, OWL-Time, and Temporal Description Logics (TDLs). Their study focused on representational efficiency, triple counts, and language support.
- Fu et al. and Hernandez et al. indirectly addressed temporal modelling through evaluations of provenance and event tracking mechanisms, emphasizing query latency and graph size implications.

Despite these contributions, temporal RDF modelling still lacks standardized evaluation methodologies. Real-world testing, semantic fidelity, and dynamic update behavior are seldom considered, making it challenging to assess model suitability in longitudinal or high-frequency data scenarios such as clinical observations or patient monitoring.

To compare different studies related to the evaluation of the temporal data modelling approaches in KGs, a comparison table was created to better understand the differences and highlight the contribution in this study. (Table 7)

Table 7- Studies related to the evaluation of temporal data modelling approaches in KGs

Study	Focus	Evaluation Metrics	Systematic methodology for designing metrics	Systematic methodology for evaluating metrics

Orlandi et al	RDF metadata representation approahces	Storage efficiency, query complexity, performance	No	No
Abuaoda et al	RDF-Star	Query execution time, storage efficiency, scalability	No	No
Wang et al.	Temporal RDF and OWL data modelling	Number of triples, query language support	No	No
Fu et al.	Evaluation of RDF provenance models	Overall graph size, query processing time	No	No
Hernandez et al.	RDF approaches for representing and querying Wiki data	Query response time	No	No
Frey et al.	Performance review of RDF metadata representation approaches	Loading time, statement count, query execution time	No	No
TEDME-KG Metrics Framework	Temporal data modelling approach	Covering Modelling. Uplifting, and Querying metrics	Yes	Yes

3.7.4 Identified limitations

A common theme that emerges from these evaluations is the narrow scope of metrics used. Many studies tend to focus on a few specific performance indicators, such as query response time, graph size, or processing time, without delving into a more comprehensive analysis of system behavior across diverse scenarios. For example, while query execution time is an important metric, it alone does not provide a complete picture of how a system handles complex query and the evaluations remain somewhat limited.

Another key limitation in many of the studies is the lack of real-world dataset evaluations. Many benchmarks are designed using synthetic or simplified datasets, which, while useful for controlled comparisons, fail to capture the nuances and challenges of real-world data environments. This gap between synthetic testing and practical application reduces the direct applicability of the findings to real-world use cases, where data is often messy, incomplete,

and distributed across different formats and sources. Without real-world testing, it is difficult to gauge how well an RDF system can handle the complexities of dynamic data, frequent updates, and varying query loads that are typical in real applications.

Moreover, the evaluations often lack a systematic methodology for defining and evaluating metrics. This absence of a more structured approach to metric design can lead to an incomplete understanding of the strengths and weaknesses of different data modelling approaches.

These findings underscore the importance of developing more nuanced evaluation frameworks that go beyond surface-level metrics and take into account the full range of system behaviors under various conditions.

To address the shortcomings identified in existing evaluation frameworks, it is clear that future benchmarking and evaluation efforts must adopt a more comprehensive and systematic approach. This would involve expanding the range of metrics used in evaluations, incorporating real-world datasets, and engaging in expert feedback to ensure that the chosen metrics reflect practical concerns in real-world applications. Such an approach would provide a more holistic understanding of how RDF systems perform across a variety of use cases, ensuring that the benchmarks are not only theoretically robust but also practically relevant.

In summary, while the existing evaluations of RDF data modelling approaches offer critical insights, they often fall short of providing a thorough and comprehensive understanding of system performance. By broadening the scope of evaluation metrics, incorporating real-world testing, and adopting a systematic methodology, future frameworks can offer more robust, reliable, and applicable benchmarks and evaluations for RDF systems in real-world KG applications.

3.8 Chapter Summary

This chapter provides a comprehensive review of the current state of research on temporal data modelling in KGs and addresses the second research objective. It presents a detailed exploration of various approaches used to model temporal data in KGs, along with an evaluation of the methodologies applied to assess these models. Through a systematic

review of the literature, this chapter identifies existing modelling approaches, highlights their strengths and weaknesses, and identifies research gaps in the field.

The chapter begins by discussing the systematic review process, which was conducted following rigorous guidelines to ensure comprehensive results. A detailed search strategy was employed across multiple databases, and the selection criteria were applied to include studies relevant to temporal data modelling in KGs. A total of 37 papers were included in the final review, with the approaches categorised based on their semantics, time dimensions, and scope of application.

A range of temporal data modelling approaches were identified, covering various applications such as spatiotemporal data, metadata modelling, and knowledge repositories. These approaches were evaluated on the basis of their use cases, including flight information, healthcare, government data, sports data, web data, and historical data. The chapter highlights the diversity in approaches, which vary from temporal-only models to more complex spatiotemporal and metadata-focused methods.

In addition to the systematic review of modelling approaches, the chapter also reviews the state of the art in evaluating RDF data modelling approaches in KGs. Several benchmarks, such as RDF Reification (REF), StarBench, and WasDiv, are introduced. However, the review notes several limitations in existing evaluations, including a narrow focus on specific performance metrics and a lack of real-world dataset testing.

The chapter identifies common limitations across the literature, such as the reliance on synthetic datasets and the absence of systematic methodologies for defining and evaluating performance metrics. These gaps present opportunities for further research and the development of more comprehensive evaluation frameworks that incorporate real-world data and expert feedback.

In conclusion, the chapter lays the groundwork for understanding the advancements and challenges in temporal data modelling and evaluation in KGs. It highlights the need for more robust, systematic evaluation methods that can provide a deeper understanding of RDF systems' performance across diverse use cases, setting the stage for the proposed research presented in subsequent chapters.

Chapter 4 Design

This chapter discusses the process of designing TEDME-KG metrics framework. The framework is used for evaluating temporal data modelling approaches in KGs. The final metric framework includes a list of metrics, metrics categories and metric subcategories, goals, questions along with their definitions, possible ranges, and units.
The chapter begins by providing an overview of the methodology for the design and evaluation of the framework, then delves deeper into the design process and finally presents the metrics framework.

4.1 Methodology Overview

The design and evaluation of TEDME-KG metrics framework were conducted in four phases. In the first phase, a literature review was performed to identify and extract metrics related to the evaluation of data modelling approaches in KGs. Following this, a thematic analysis was conducted, leading to the development of the first iteration of the metrics framework. In the second phase, an experiment was conducted as a pilot study. The metrics framework was refined during the design of two knowledge graphs based on two selected temporal data modelling approaches using the FAIRVASC project case study. During the KG design process, the first iteration of the metrics framework was reviewed using the Goal Question Metrics (GQM) approach [33, 131]. Following this systematic process, the second iteration of the metrics framework was designed.

In the third phase, the metrics framework was evaluated based on experts' feedback. A survey was conducted to assess the appropriateness of the metrics framework according to a set of criteria. Experts' suggestions on metrics, grouping, goals, and questions were solicited, and the general usability and potential usefulness of the framework were investigated through a sample scenario.

In the fourth phase, an experimental study was performed. Temporal data modelling approaches were evaluated based on various temporal questions, and the applicability of the framework was assessed using the FAIRVASC use case.

Phase 1 and 2 are discussed in this chapter. Phase 3 will be discussed in chapter 5 and phase 4 will be discussed in chapter 5. (Figure 10)

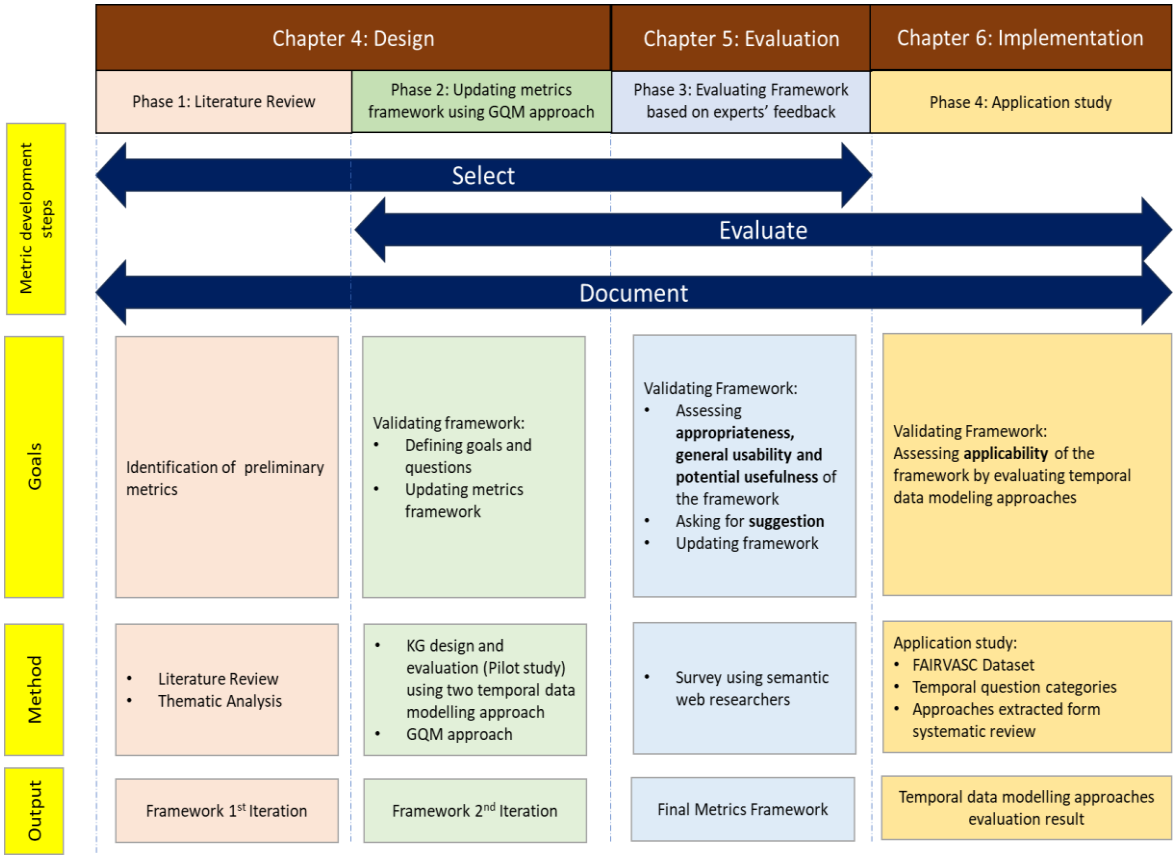


Figure 10- Design and evaluation phases of TEDME-KG metrics framework

The detailed process is discussed in detail below:

4.2 Phase 1: Literature Review

In this phase, metrics were identified in a three-step process. Initially, literature reviews were conducted, and metrics were selected from three diverse sources. Three different literature reviews were conducted, and various metrics were extracted from the three different sources. The sources include the following:

- Source 1 (S1): Literature review on metrics for evaluating data modelling approaches in KGs
- Source 2 (s2): Literature review on survey (review) papers about temporal data modelling approaches in KGs
- Source 3 (S3): Systematic review on temporal data modelling approaches in KGs

Following this, a thematic analysis was performed, leading to the assignment of a theme to the group of relevant metrics and the first iteration of metrics was developed. This included a comprehensive table that detailed the metrics, their associated parent themes and sub

themes, units of measurement, possible answers/ranges, and definitions. The detail of each step is discussed below:

The literature reviews details are discussed below:

4.2.1 Literature review on metrics for evaluating data modelling approaches in KGs – (S1)

A literature review was performed to identify and extract metrics relating to the evaluation of RDF data modelling approaches. The scope of the search extended beyond just metrics for temporal data modelling, due to the scarcity of relevant papers in this area. By broadening the search criteria, it was possible to include all studies related to the evaluation of data modelling approaches. The method of the literature review is discussed below:

4.2.1.1 Formulating The Research Question

The literature review was conducted using the CIMO (Context, Intervention, Mechanism, and Outcome) question framework for developing the research question as follows:

- Context: RDF
- Intervention: Evaluation
- Mechanisms: metric
- Outcome: modelling

The research question was: what are the metrics related to the evaluation of data modelling approaches in RDF?

4.2.1.2 Information Sources

The Scopus, ACM Digital library, and IEEE Xplore databases were searched for the literature review.

4.2.1.3 Search Strategy

The search concepts and keywords based on CIMO framework are set out in Table 8.

The search query used for the databases was:

“(rdf OR "resource description framework") AND (evaluation OR assessment OR assess OR evaluate) AND (Criteria OR criterion OR metric OR measure OR attribute) AND (representation OR model OR design) “

Table 8- Search concepts and keywords for use in a literature review to identify metrics relating to the evaluation of RDF data modelling approaches

CIMO Framework	Concept	Keywords
Context	RDF	RDF, resource description framework
Intervention	Evaluation	Evaluation, evaluate, assess, assessment
Mechanism	Metric	Metric, Criteria, criterion, measure, attribute
Outcome	Modelling	Model, Representation, design

4.2.2 Survey papers on temporal data modelling approaches in KGs - (S2)

A hand searching was conducted to identify surveys related to temporal data modelling approaches, meta data modelling approaches. Google scholar was searched to identify survey papers/review papers/state of art review papers. Nine papers that reviewed scientific literature on addressing temporal data modelling approaches, meta data modelling in KGs were identified. (Appendix C)

4.2.3 Systematic review on temporal data modelling in KGs – (S3)

33 papers which were identified in a systematic review to identify temporal data modelling approaches in KGs were assessed. The detail of this review has been discussed in chapter 3. Final 33 papers identified in this review were included with the aim of extracting metrics. (Appendix C)

4.2.4 Data extraction and synthesis process

Retrieved literature were imported into Covidence, systematic review software, for screening. [49] Firstly, titles and abstracts were reviewed for inclusion. The final selection of papers was determined by doing a full-text review of the screened literature. Information from each included paper was entered into a data extraction form that was incorporated into Covidence. A single researcher thoroughly reviewed the publications before manually entering the data into the data extraction table. Records of the reference and metrics were made. In order to do analysis, the data extraction form was then exported into a spreadsheet as a table. (Appendix C)

4.2.5 Literature review results

19 papers published between 2009 and 2021 were identified for inclusion. While the 19 papers included in the review span over a decade (2009–2021) and were selected using a structured and comprehensive search strategy across major academic databases (Scopus, ACM Digital Library, and IEEE Xplore), it is acknowledged that this selection may not fully represent the entire universe of available evaluation metrics for RDF data models. To mitigate this, an additional search was conducted in 2024 using the same strategy to identify any newly proposed or overlooked metrics; however, no new metrics were identified that had not already been captured in the original review.

The papers were assessed for their credibility, methodological rigor, relevance, and impact. Most papers were published in peer-reviewed journals or conferences, ensuring their academic credibility. Methodologies in these papers were generally well-defined and transparent, often including empirical evaluations, benchmark comparisons, or structured surveys. The selected sources were highly relevant to the research topic, focusing on RDF data modelling, assessment, and temporal knowledge graph evaluation

4.2.6 Thematic analysis

4.2.6.1 Methodology

An inductive thematic analysis of the retrieved literature was conducted to categorise metrics related to the evaluation of data modelling approaches in RDF [132]. The thematic analysis was conducted in two steps:

Step 1: A theme was assigned to each extracted item based on their concept. The metric themes were not unique per item. For example, where 2 items referred to the same concept, the theme was repeated for both.

Step 2: The themes were reviewed and agreed by a second researcher.

Once the metric list had been compiled, thematic statements for the metric categories were formulated, and definitions, possible ranges, and units for each metric were defined.

4.2.7 Thematic Analysis results

A total of 49 metrics were extracted from the literature. Removal of duplicates and matching the metric names, resulted in 16 unique metrics. The 16 identified metrics were categorised in three sub themes namely, implementation support and standardisation, complexity, and performance. Metric subcategories were then assigned to three parent themes.

namely, modelling, uplifting, and querying. The sub themes were named metrics subcategories, and the parent themes were named metrics categories. The three main categories—Modelling, Uplifting, and Querying—emerged during the thematic analysis conducted as part of developing the evaluation framework. During this process, individual metrics were identified across various approaches and grouped based on their conceptual similarity and role in the data lifecycle.

- Modelling includes metrics related to how temporal information is represented in RDF, focusing on structural characteristics and compliance with standards.
- Uplifting captures the process of transforming raw data into RDF, emphasizing transformation complexity, tool support, and data expansion.
- Querying covers the practical aspects of retrieving temporal data from RDF graphs, highlighting aspects of query execution and system responsiveness.

These themes reflect distinct but interconnected stages in the use of RDF for temporal data and help organize the evaluation in a meaningful, task-oriented way.

The sources from which the quality metrics were reused were assessed in terms of their academic rigour, publication venue, and relevance to RDF data modelling. The majority of the 19 included papers were peer-reviewed and published in reputable journals and conferences within the semantic web, knowledge representation, and data engineering domains. These venues are recognized for their methodological soundness and impact in the field. Most studies employed empirical methods, including benchmarking, prototype implementation, and comparative analysis, enhancing the reliability of the metrics they proposed. While some papers had a narrower focus (e.g., performance only), they contributed specific, clearly defined metrics grounded in practical use cases. The thematic synthesis allowed triangulation across multiple sources, reinforcing the validity of the identified metrics despite individual limitations in scope or evaluation methodology.

Thematic statements for three metrics categories are presented in Table 9.

Table 9- Thematic statements for the evaluation phases

Parent theme (Metrics category)	Thematic statement

Modelling	This category refers to modelling temporal data in RDF. This category contains two metric subcategories including implementation support and standardisation, and complexity. The implementation support and standardisation is concerned with the implementation support tool and adherence of the data modelling approaches to W3C standards. Modelling complexity is measured by the graph size, number of statements, resource redundancy, and additional triple generation. More detail on the final metrics definitions- see Table 13
Uplifting	This category refers to transforming data to RDF. This metrics category contains three metrics subcategories including implementation support and standardisation, complexity, and performance. The implementation support and standardisation are concerned with the implementation support tool and adherence to W3C standards. The complexity of the uplifting is concerned with the number of triple maps and term maps used in the uplift code. The performance subcategory refers to the storage space. More detail on the final metrics definitions- see Table 13
Querying	This theme refers to querying RDF data based on temporal question categories in a graph database. This metrics category contains two metric subcategories including implementation support and standardisation, and performance. More detail on the final metrics definitions- see Table 13

4.2.8 Metrics framework- First iteration

The initial version of the metrics framework was developed based on findings from the literature reviews and the results of the thematic analysis.

Table 10- Metrics framework for evaluating temporal data modelling approaches in KGs – First Iteration

Metrics category	Metrics subcategory	Metrics
Modelling	1. Implementation support and standardisation	Implementation support tool
		Adherence to W3C standards
	2. Complexity	Additional triple generation
		Number of nodes
		Number of triples
		Resource redundancy
Uplifting	1. Implementation support and standardisation	Implementation support tool
		Adherence to W3C standards
	2.Complexity	Number of triple maps
		Number of term maps
	3. Performance	Storage space
Querying	1. Implementation support and standardisation	SPARQL query language support
		Triple store support tool
	2. Performance	Query length requirement to execute a particular task
		Data load time
		Query response time

4.3 Phase 2: Updating metrics framework through GQM approach.

An experiment was conducted as a pilot study to evaluate two different temporal data modelling approaches and to update the TEDEME-KG metrics framework. The experiment concerns a use case based on the FAIRVASC project, which was discussed in section 2.

4.3.1 Pilot Study

The temporal data modelling approaches used in this experiment, were "Standard Reification" and "Named Graph" [108, 109]. The Standard Reification approach enables the expression of the temporal dimension of RDF triples using RDF through two main steps. First, a triple is represented by a Statement instance, with the subject, predicate, and object indicated separately in three distinct triples. Second, assertions are made about that instance as if it were a statement. Named Graphs facilitate the assignment of a URI to a collection of triples, allowing for statements to be made about the entire set.

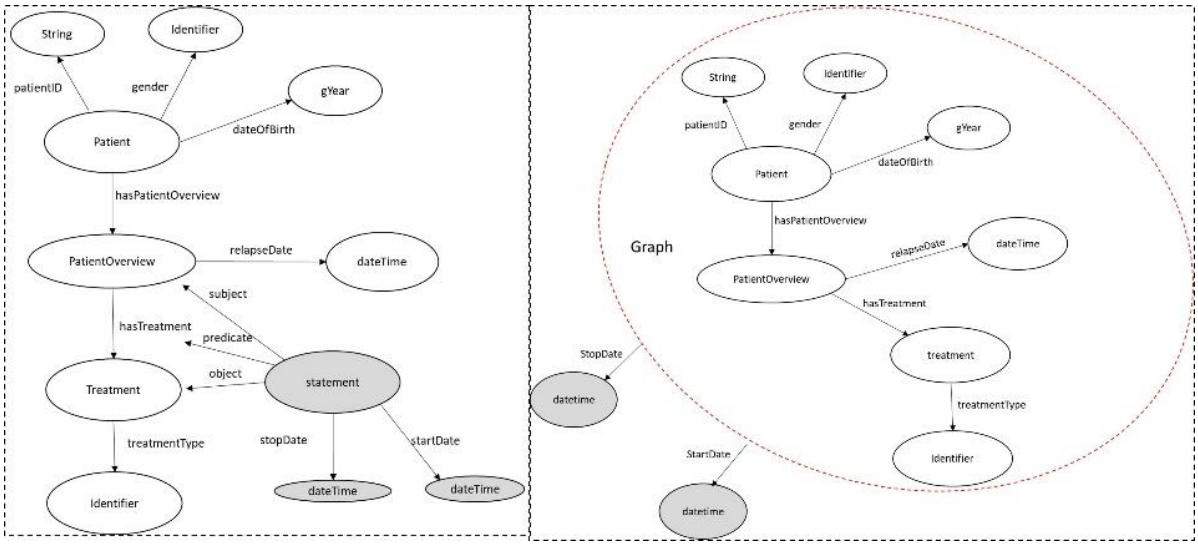


Figure 11- Adding medication time to the base model based on Standard Reification approach (Left side) and Named Graph approach (Right side)

Firstly, two ontologies were designed based on these two approaches, incorporating medication start time and stop date into the base model. (Figure 11) The Base model was discussed in chapter 2. (Figure 6)

Once the ontologies were created, R2RML codes were developed for each of the two approaches. The data were uplifted using the R2RML-F[4] uplift engine and RDF data were generated for each approach. The two RDF datasets were loaded on a GraphDB[5] version 10.3.1 triple store running on a single pc with a 4-cores CPU and 32GB of main memory.

[4] https://github.com/chrdebru/r2rml
[5] https://graphdb.ontotext.com/

A competency question was used to test the metrics in “Querying” category. The competency questions guide the structure and objectives of the knowledge graph, ensuring it meets the needs of the use case. The competency questions are discussed below.

4.3.1.1 Competency Questions

Based on the existing literature, temporal questions can be categorised into 4 different categories including explicit temporal, implicit temporal, ordinal temporal, and questions with temporal answers [133, 134]. The description of each category is outlined below:

- Explicit temporal: The time is explicitly mentioned in the question.
- Implicit temporal: The time is implicitly mentioned in the question.
- Ordinal temporal: The question contains ordinal numbers such as, first, second etc.
- Temporal answers: The answer to these questions contains time related data such as, day, month, year etc.

An implicit temporal question was considered for this experiment, “Which medications used by a patient with ID=12, one month before and one month after his/her relapse?”

Two SPARQL queries were created based on each approach. The queries were executed, and the KG modelling approaches were evaluated based on the defined metrics. The SPARQL queries, and uplift codes (extended parts) are available on GitHub repository.[6] It should be noted that given the confidentiality of medical data, the dataset was not uploaded in the repository.

During the KG design process, a goal was defined at each step considering the metrics categories. Based on these goals, relevant questions were formulated and assigned to the appropriate metric subcategories. The metrics were then reviewed, and the following changes were applied.

4.3.2 Framework Refinement

Five changes were made to the framework's first iteration.

Change 1: In the modelling category, "Implementation support tool" was changed to "Ontology editor support." This adjustment made the metric more explicit, as the ontology editor tool was the only implementation support tool identified at this design stage.

[6] https://github.com/sepidha/PhD/tree/KGSWC-2024

Change 2: In the modelling category, "Adherence to W3C" was changed to "Ontology adherence to W3C." This adjustment made the metric more explicit, as ontology adherence to W3C was the only form of adherence identified at this design stage.

Change 3: In the Uplifting category, "Implementation support tool" was changed to "Uplift engine support." This adjustment made the metric more explicit, as the uplift engine support was the only implementation support tool identified at this design stage.

Change 4: In the Uplifting category, "Adherence to W3C" was changed to "Uplift adherence to W3C." This adjustment made the metric more explicit, as uplift adherence to W3C was the only form of adherence identified at this design stage.

Change 5: In the Uplifting category, "storage space" was revised to "RDF data storage space." This adjustment clarified the metric.

Based on the specified changes and updates, the second version of the TEDME-KG metrics framework was developed (Table 11). In this version, goals and questions were incorporated to enhance clarity and aid in the evaluation of temporal data modelling approaches. The modifications in the metrics are highlighted in bold.

Table 11- TEDME-KG Metrics framework -2nd iteration

Goal	Metrics Category	Metrics Subcategory	Question	Metric
Evaluating temporal data modelling approaches in terms of modelling aspect	Modelling	Implementation support and Standardisation	Are the approaches supported by an ontology editor?	**Ontology editor support**
			Do the approaches use W3C recommendations?	**Ontology adherence to W3C**
		Complexity	What is the number of nodes?	Number of nodes
			What is the number of triples?	Number of triples
			What is the number of additional generated triples?	Additional triple generation

			What is the number of redundant resources?	Resource redundancy
Evaluating temporal data modelling approaches in terms of uplifting aspect	Uplifting	Implementation support and Standardisation	Is there any uplift engine available for uplifting data based on the modelling approach?	**Uplift engine support**
			Does the uplifting use W3C recommendation?	**Uplift adherence to W3C**
		Complexity	What is the number of triple maps?	Number of triple maps
			What is the number of term maps?	Number of term maps
		Performance	What is the storage space after uplifting data?	**RDF Data storage space**
Evaluating temporal data modelling approaches in terms of querying aspect	Querying	Implementation support and Standardisation	Are the approaches supported by a W3C query language?	W3C Query language support
			Are these approaches supported by a triple store?	Triple store support
		Performance	What is the data load time in the triple store?	Data load time
			What is the query length?	Query length
			What is the query response time?	Query response time.

4.3.3 Experiment Results

This section presents the results of evaluating Standard Reification and Named Graph. Using the TEDME-KG metrics framework, the evaluation focused on three main aspects: modelling, uplifting, and querying. Below are the detailed findings for each category.

4.3.3.1 Results in modelling metrics category

Neither the standard reification approach nor the named graph approach supports ontology editor tools, indicating a potential gap in user-friendly tools for ontology design. However, both approaches adhere to W3C standards, ensuring compatibility with established Semantic Web technologies. In terms of complexity, the standard reification approach has more nodes and triples, indicating a more granular or fragmented representation of data and more detailed relationships, which adds complexity. It also generates a higher number of additional triples and shows higher resource redundancy, suggesting more duplicated resources and potential inefficiency.

4.3.3.2 Results in uplifting metrics category

Both approaches support uplift engines, facilitating data transformation, and adhere to W3C standards during data uplifting, ensuring consistency with web standards. The standard reification approach requires more triple maps and term maps, indicating a more complex mapping process and higher complexity in the data transformation rules.

4.3.3.3 Results in Querying metrics category

Both approaches support W3C query languages, ensuring compatibility with standard querying practices, and are supported by triple stores, facilitating their practical use in data management systems. In terms of performance, the named graph approach has a significantly faster data load time and query response time, indicating better performance in data ingestion and query execution. Both approaches use the same amount of disk space for queries, suggesting similar storage requirements for querying purposes, but the named graph approach has a slightly shorter query length, indicating marginally more concise query formulation.

4.3.3.4 Summary

The standard reification approach is more complex, with more nodes, triples, additional triples, triple maps, and term maps. The named graph approach outperforms the standard reification approach in terms of storage space, data load time, and query response time. Named graph shows better efficiency with less resource redundancy and lower storage requirements. Both approaches adhere to W3C standards and support necessary tools for uplift and querying. If detailed and rich data representation is needed, the standard reification approach might be suitable despite its complexity. However, if efficiency and performance are prioritised, the named graph approach is the better choice.

The numbers for each metric for the two approaches, Standard Reification and Named Graph, are provided in Table 12.

Table 12- Evaluation Results of Named Graph and Standard Reification based on TEDME-KG Metrics Framework

Metrics Category	Metrics Subcategory	Metric	Approach 1: Standard Reification	Approach 2: Named Graph
Modelling	Implementation support and Standardisation	Ontology editor support	No	No
		Ontology adherence to W3C	Yes	Yes
	Complexity	Number of nodes	11636	9345
		Number of triples	43348	28618
		Additional triple generation	34414	19684
		Resource redundancy	42134	25113
Uplifting	Implementation support and Standardisation	Uplift engine support	Yes	Yes
		Uplift adherence to W3C	Yes	Yes
	Complexity	Number of triple maps	5	3
		Number of term maps	17	12
	Performance	RDF Data storage space	5125	3387
Querying	Implementation support and Standardisation	W3C Query language support	Yes	Yes
		Triple store support	Yes	Yes
	Performance	Data load time	2000	1000
		Query length	20	18
		Query response time.	400	100

4.3.4 Detailed description of the metrics

Once the metrics framework was updated a detail description of each metric, metrics category and metrics subcategories were prepared. This section provides an overview of the 16 metrics utilised in the TEDME-KG Metrics Framework (Table 13) along with the metrics categories and subcategories. (Table 14) Each metric is detailed with the following structure:

- **Definition**: A clear explanation of the metric
- **Range/Answer**: The possible values or answers for the metric.
- **Unit**: The unit of measurement used, if applicable.

Table 13- Metrics details used in TEDME-KG Metrics framework.

No	Metric	Definition	Range/Answer	Unit
1	Ontology editor support	This metric shows whether an approach is supported by a tool for ontology development.	Yes/No	Not Applicable
2	Ontology adherence to W3C	This metric shows whether an approach uses the W3C recommendations.	Yes/No	Not Applicable
3	Number of nodes	This metric shows the number of nodes used for modelling.	0-n	Node
4	Number of triples	This metric shows the number of triples used for modelling.	0-n	Triple
5	Additional triple generation	This metric shows the number of additional triples generated when adding time to the model. (Number of triples after adding time aspect –	0-n	Triple

		Number of triples before adding time aspect)		
6	Resource redundancy	This metric shows whether the existing approach generates redundant resources or not. (Number of triples and nodes after adding time aspect) - (number of triples and nodes before adding time aspect)	0-n	Triple and Node
7	Uplift engine support	This metric shows whether an approach is supported by an uplifting engine.	Yes/No	Not Applicable
8	Uplift adherence to W3C	This metrics shows whether the uplifting uses a W3C recommendation.	Yes/No	Not Applicable
9	Number of triple maps	This metric shows the number of triple maps used in the uplift code. A triple map defines the logical table, which may be a table or view from a relational database.	0-n	Triple maps
10	Number of term maps	This metric shows the number of term maps in the uplift code. The subject map, the predicate map and the object map are called term maps.	0-n	Term maps

11	RDF Data storage space	This metric shows the size of the RDF dataset generated by the approach after uplifting data.	0-n	KB
12	W3C Query language support	This metric shows whether an approach is supported by a W3C query language recommendation.	Yes/No	Not Applicable
13	Triple store support	This metric shows whether an approach is supported by a triple store.	Yes/No	Not Applicable
14	Data load time	This metric shows the time in which the RDF data is being uploaded in the triple store.	0-n	second
15	Query length	This metric shows the query length required to retrieve information to execute a particular task.	0-n	Line
16	Query response time.	This metric shows the time in which the query is being executed.	0-n	second

Table 14- Metrics Categories and subcategories definitions

Metrics Category	**Definition**	**Metrics Subcategory**	**Definition**
Modelling	This category refers to modelling temporal data in RDF. This category contains two	Implementation support and Standardisation	The implementation support and standardisation is concerned with the

	metric subcategories including implementation support and standardisation, and complexity.		implementation support tool and adherence of the data modelling approaches to W3C standards.
		Complexity	Modelling complexity is measured by the graph size, number of statements, resource redundancy, and additional triple generation
Uplifting	This category refers to transforming data to RDF. This metrics category contains three metrics subcategories including implementation support and standardisation, complexity, and performance	Implementation support and Standardisation	The implementation support and standardisation are concerned with the implementation support tool and adherence to W3C standards.
		Complexity	The complexity of the uplifting is concerned with the number of triple maps and term maps used in the uplift code
		Performance	The performance subcategory refers to the storage space.

Querying	This theme refers to querying RDF data based on temporal question categories in a graph database. This metrics category contains two metric subcategories including implementation support and standardisation, and performance	Implementation support and Standardisation	The implementation support and standardisation is concerned with the conformance of the approach to SPARQL query language and the availability of a triple store to support the modelling approach.
		Performance	Performance is concerned with the required query length to execute a particular task, data load time, and query response time

Below is a summary of each category and the value of its associated metrics.

Modelling

- Implementation Support and Standardisation

  Why it matters: Standards-compliant and tool-supported models are easier to adopt, integrate, and scale in real-world Semantic Web applications.

- Complexity

  Why it matters: High complexity can lead to maintenance challenges, increased storage, and reduced query performance. Simpler models are generally more efficient and interpretable.

Uplifting

- Implementation Support and Standardisation

Why it matters: Standards-based uplifting ensures portability, reduces custom coding, and enables interoperability with existing tools.

- Complexity
  Why it matters: More mappings often indicate more effort, higher error potential, and lower maintainability.

- Performance (Storage Space)

  Why it matters: Larger output can increase storage and computational requirements; efficient representation is crucial for scalable systems.

Querying

- Implementation Support and Standardisation

  Why it matters: Without SPARQL compatibility or triple store support, querying becomes impractical and limits real-world application.

- Performance
  Why it matters: These are core indicators of how efficiently a model supports real-time querying and user interaction, especially in time-sensitive or large-scale environments.

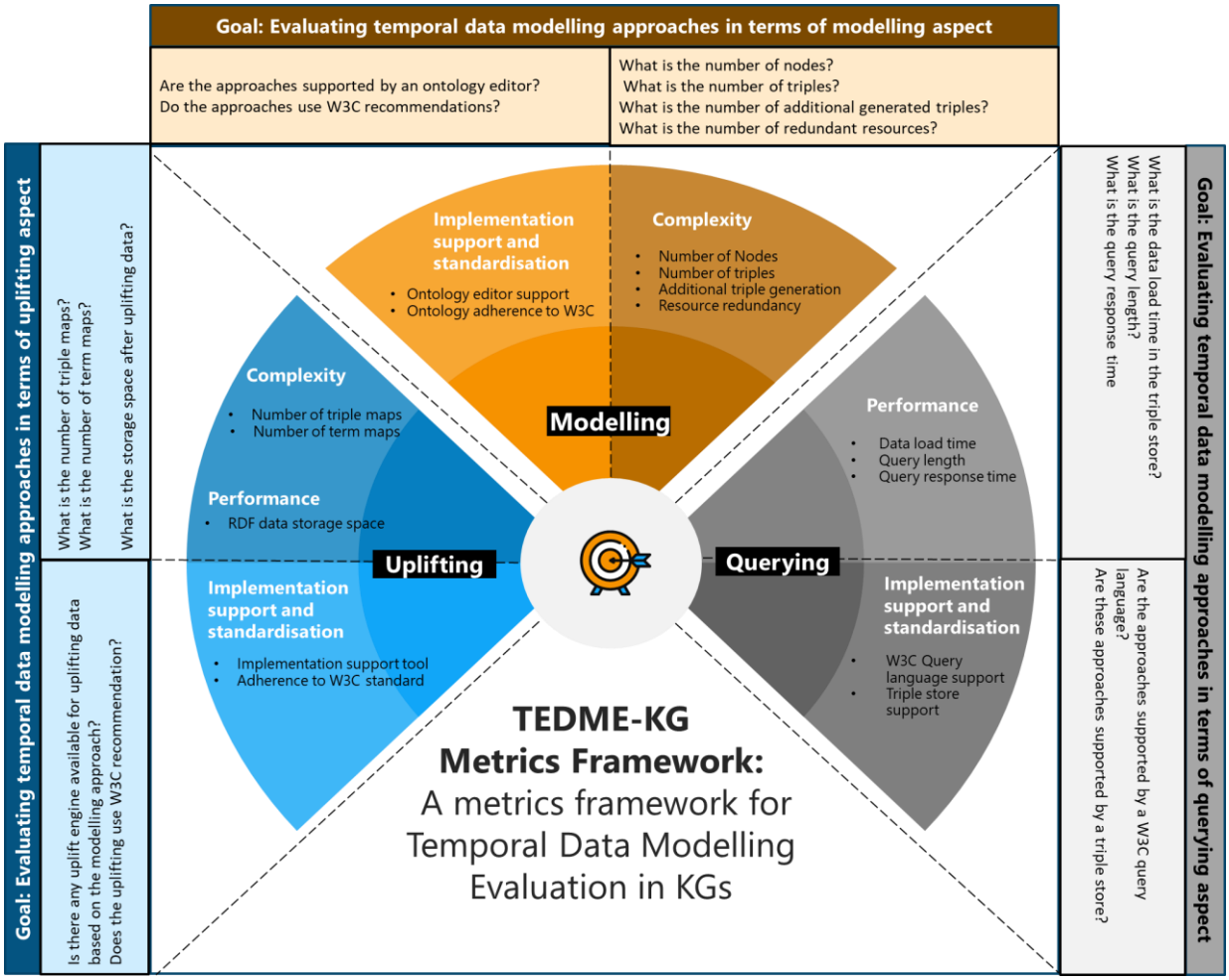
While the TEDMEKG metrics framework was developed specifically to evaluate temporal data modelling approaches in Knowledge Graphs (KGs), it shares conceptual similarities with the Linked Data Quality Assessment framework proposed by Zaveri et al. Both frameworks focus on structured evaluation of datasets or modelling methods, particularly concerning aspects such as standardisation, tool support, and performance. For instance, TEDMEKG's metrics like Ontology adherence to W3C, Triple store support, and W3C Query Language support align with Zaveri et al.'s dimensions of interoperability and accessibility, which emphasize the use of common standards and protocols. Similarly, TEDMEKG's attention to query response time and data load time touches on performance and efficiency, which are also discussed in the Linked Data Quality framework. However, TEDMEKG introduces additional metrics tailored to the temporal dimension—such as additional triple generation, resource redundancy, and number of term maps—which are not explicitly covered in Zaveri's model. Thus, while there is overlap in foundational quality dimensions, TEDMEKG extends the discussion by introducing metrics specifically suited

for temporal semantics in KGs, filling a niche not addressed in existing Linked Data quality frameworks [122].

4.4 Visual representation of TEDME_KG Metrics Framework

A detailed visual representation of the TEDME-KG metrics framework was designed to enhance understanding. This visual model captures the core components of the framework, illustrating how various metrics are categorised to evaluate temporal data modelling approaches in knowledge graphs.

Figure 12- Visual representation of TEDME-KG Metrics Framework



4.5 Chapter Summary

This chapter outlines the design and development of the TEDME-KG metrics framework, which is used to evaluate temporal data modelling approaches in knowledge graphs (KGs). The framework includes a list of metrics, categorised into various subcategories, along with their definitions, possible ranges, and units. The chapter provides a comprehensive look into the methodology, focusing on the first two phases of a four-phase process. Phases 3 and 4 are left for discussion in the following chapter.

The first phase began with a literature review to extract relevant metrics. Three distinct reviews were conducted: one on metrics for evaluating data modelling in KGs, one on surveys related to temporal data modelling, and a systematic review of temporal data modelling approaches. From these reviews, 42 metrics were identified, narrowed down to 16 unique ones through a thematic analysis, which categorised the metrics into modelling, uplifting, and querying phases.

The second phase involved refining the framework by applying it to the design of two knowledge graphs, based on the FAIRVASC project, using two temporal data modelling approaches: Standard Reification and Named Graph. The framework was reviewed using the Goal Question Metrics (GQM) approach, which helped clarify and refine metrics related to the design process. Six major changes were made to the first iteration of the framework, enhancing clarity and ensuring better evaluation alignment.

The chapter concludes by presenting the results of the pilot study, comparing the two modelling approaches using the updated metrics framework. Key findings revealed that the Named Graph approach was more efficient in terms of storage space, data load time, and query response time, while the Standard Reification approach demonstrated higher complexity but offered more detailed data representation. The updated metrics framework, now in its second iteration, provides a robust tool for evaluating temporal data modelling in KGs.

The TEDME-KG metrics framework does not aim to identify the best approach among temporal data modelling approaches in KGs. Instead, the proposed framework is intended to assist Semantic Web researchers by highlighting various aspects to consider when modelling temporal data in KGs. Based on the defined metrics, categories, and subcategories, researchers can determine which approach best aligns with their specific use

case and requirements. The framework deliberately avoids aggregating all metrics into a single final score, as the relative importance of each metric can vary significantly depending on the context. For instance, one use case may prioritise query performance, while another may value modelling simplicity. Combining metrics into an overall score would risk oversimplifying these trade-offs and obscuring the specific strengths and weaknesses of each modelling approach. Instead, the framework supports a comparative, multidimensional evaluation, enabling users to make informed decisions based on the metrics most relevant to their particular domain, application, and constraints.

Overall, this chapter lays the foundation for the further refinement and application of the TEDME-KG framework, which will be elaborated in the subsequent phases and chapters

Chapter 5 Evaluation

In this chapter TEDME-KG metrics framework was evaluated using the experts' feedback. The appropriateness, and general usability of the TEDME-KG metrics framework were assessed using experts' feedback. This section discusses the evaluations and their results.

5.1 Strategy

An anonymous survey was conducted, and it was circulated through the Researcher's supervisor network to experts working in the area of Semantic Web. The survey included three sections. In the first section the appropriateness of the metrics framework for evaluating temporal data modelling approaches were evaluated. In section 2 experts' suggestion regarding the metrics framework were assessed. Finally, in the third section the general usability of the framework was evaluated. (Appendix C)

5.2 Sample Size

The sample size of the experiment is outlined below.

- **Background**: Participants which are called experts are Semantic Web researchers who are very knowledgeable with Semantic Web technologies, with at least one year of hands-on experience in knowledge graph design, ontology development, R2RML and SPARQL. The participants have experience in developing ontologies in a research environment.
- **Number of Participants**: The study involved 10 expert participants.

5.3 Evaluation of appropriateness of the TEDME-KG metrics framework

This section discusses the evaluation of appropriateness of the TEDME-KG metrics framework based on a set of criteria. The criteria established by Keeney et al [27] are highly relevant to the evaluation of the TEDME-KG metrics framework. These criteria ensure that the framework is not only theoretically sound but also practically applicable, making it a reliable tool for researchers and practitioners in the field of Knowledge Graphs. Below is an extended discussion on how these criteria apply to the TEDME-KG metrics framework. The criteria used in assessing appropriateness include:

Relevant: The metrics framework is related to the fulfilment of objective.

Unambiguous: The metrics framework is presented in a clear way that is obvious what is being measured.

Direct: The metrics framework addresses objectives clearly.

Operational: The metrics framework is practical and feasible.

Understandable: The metrics framework is presented in a way that is easy to interpret.

Comprehensive: The metrics framework thoroughly covers all of the defined objectives

5.3.1 Results and interpretation

Ten Semantic Web researchers participated in the survey, including six PhD Researchers and four postdoctoral researchers. Table 15 presents the distribution of participants based on years of experience.

Table 15- Distribution of participants in the survey based on years of experience.

Years of Experience	Number of participants
1 < years <=2	3
2 < years <=5	6
5 < years	1

The respondents provided their insights on the appropriateness of the framework for evaluating temporal data modelling approaches in knowledge graphs (KGs). (Figure 13)

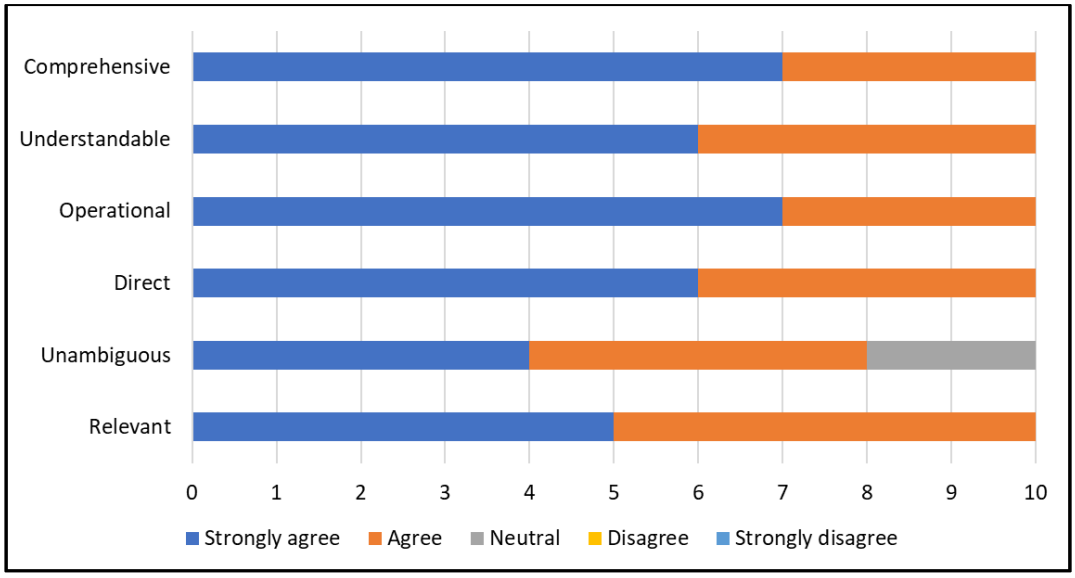


Figure 13- Evaluation of the appropriateness of the TEDME-KG metrics Framework based on a set of criteria.

Below is a detailed interpretation and analysis of the results.

5.3.1.1 Relevance

The feedback on the relevance of the metrics framework was positive, with all experts either strongly agreeing (5 experts) or agreeing (5 experts) that the framework is aligned with its intended objectives. This indicates that the framework relates to the fulfilment of the objectives.

5.3.1.2 Unambiguity

While the majority of experts found the framework to be unambiguous, as indicated by the combined total of 8 experts who either strongly agreed or agreed, there were 2 experts who provided a neutral response. This neutrality suggests that some aspects of the framework might not be entirely clear or could be open to multiple interpretations. The feedback implies that while the framework is generally clear, there may be specific areas where the definitions or presentations of metrics could be refined to reduce any potential ambiguity.

5.3.1.3 Directness

The directness of the framework—how clearly it addresses its objectives—was highly rated, with 6 experts strongly agreeing and 4 agreeing. This strong positive response indicates that the framework's metrics are well-aligned with its goals and that users can see how the metrics support the intended outcomes.

5.3.1.4 Operational Feasibility

The operational feasibility of the framework received the highest number of strong agreements (7 experts), with the remaining 3 experts agreeing. This suggests that the framework is not only theoretically sound but also practical and feasible to implement in real-world scenarios. The absence of any neutral or negative responses indicates a strong consensus that the framework is designed with practical usability in mind. This is a critical factor for ensuring that the framework can be adopted and used by practitioners and researchers in the field.

5.3.1.5 Understandability

Understandability is another area where the framework scored highly, with 6 experts strongly agreeing and 4 agreeing that the metrics are easy to interpret. This feedback

indicates that the framework is accessible and that users can easily grasp its structure and content.

5.3.1.6 Comprehensiveness

The high level of strong agreement (7 experts) and agreement (3 experts) suggests that the framework is viewed as comprehensive, thoroughly covering all necessary aspects. The lack of neutral or negative feedback underscores the framework's adequacy in addressing the full scope of relevant issues, indicating that it successfully captures the complexity of the domain it is designed to measure.

5.3.2 Overall Analysis of appropriateness

Overall, the TEDME-KG Metrics Framework is considered highly appropriate by the experts, with strong agreement across all evaluated criteria. The framework is seen as relevant, direct, operationally feasible, understandable, and comprehensive. The main area for potential enhancement lies in improving clarity to eliminate any remaining ambiguities.

5.4 Validation of TEDME-KG metrics framework design

This section delves into the validation process for the TEDME-KG metrics framework, highlighting the feedback and insights gained from expert participants. The validation process is critical to ensuring that the framework is robust, comprehensive, and adaptable to the needs of its users. The section begins with an overview of the evaluation structure, followed by a detailed discussion of the results for each ontology. This approach ensures that the design of the TEDME-KG metrics framework is thoroughly examined and refined based on expert input.

The evaluation focused on three key questions designed to gather detailed feedback on the design and structure of the TEDME-KG metrics framework. These questions were aimed at uncovering any potential gaps, identifying areas for improvement, and ensuring the framework's metrics are both comprehensive and clearly organised. The specific questions posed to the participants were:

5.4.1 Suggestions for New Goals and Questions

Question*:* *"Do you have any suggestions for adding a new goal and/or question in any of the metric categories or metric subcategories or changing the existing ones?"*

The first question asked participants whether they had any suggestions for adding new goals or questions within any of the existing metric categories or subcategories. Additionally, participants were encouraged to provide input on whether the current goals and questions should be modified. This question is crucial for understanding whether the framework's current goals are exhaustive and if the questions effectively capture the necessary aspects of temporal data modelling in Knowledge Graphs. The feedback gathered here helps in identifying potential oversights and ensuring that the framework remains aligned with evolving research needs.

5.4.2 Suggestions for New Metrics, Metrics Categories, or Metrics Subcategories

Question: *"Do you have any suggestions for adding a new metric, metric category, or metric subcategory or changing the existing ones?"*

The second question sought to explore whether there were any recommendations for introducing new metrics, metric categories, or metric subcategories, or whether any existing ones should be altered. This question is particularly important for refining the granularity and scope of the framework. By asking this, the evaluation aimed to ensure that the metrics are not only comprehensive but also detailed enough to capture the nuances of temporal data in Knowledge Graphs. The responses to this question are expected to provide insights into whether the current metrics are sufficient or if additional metrics are needed to cover other critical dimensions.

5.4.3 Suggestions for Grouping and Labelling of Metrics

Question*:* *"Do you have any suggestions for grouping or labeling the metrics, metric categories, or metric subcategories?"*

The third question focused on the organisational aspects of the framework, specifically the grouping and labeling of metrics, metric categories, and subcategories. Participants were asked to provide suggestions on whether the existing structure was logical and intuitive or if it could benefit from reorganisation. This feedback is essential for ensuring that the framework is user-friendly and that the metrics are presented in a way that facilitates easy navigation and understanding. Effective grouping and labeling are crucial for the practical application of the framework, as they help users quickly locate relevant metrics and understand their relationships.

Through these questions, the validation process aimed to capture a broad range of feedback, ensuring that the TEDME-KG metrics framework is not only scientifically rigorous but also practically applicable and responsive to the needs of its users. The subsequent sections will discuss the feedback received from the participants in detail, providing a comprehensive analysis of the suggested improvements and their implications for the future development of the framework.

5.4.4 Results and interpretation

The feedback provided by participants on the TEDME-KG metrics framework is generally positive, reflecting a strong alignment between the framework's design and the needs and expectations of its users. The comments received have been categorised into three main areas: suggestions for new goals and questions, suggestions for new metrics or categories, and suggestions for grouping and labeling. Below is an analysis of the overall feedback in each area.

5.4.4.1 Suggestions for New Goals and Questions

No participants suggested the addition of new goals or questions, nor did they recommend modifications to the existing ones. This lack of suggestions could be interpreted as a strong endorsement of the current framework's comprehensiveness. The feedback received includes a comment stating that:

"The goals and questions in the framework are comprehensive and cover the aspects I would expect."

This indicates that participants generally feel the framework is thorough and well-rounded, effectively addressing all necessary aspects of temporal data modelling in Knowledge Graphs. The absence of suggestions for new goals or modifications suggests that the

framework’s current objectives and questions are perceived as sufficiently robust and relevant to the domain.

5.4.4.2 Suggestions for New Metrics, Metrics Categories, or Metrics Subcategories

Similarly, there were no suggestions for the addition of new metrics, metric categories, or subcategories, nor for the modification of the existing ones. This indicates a general satisfaction with the current structure and content of the framework. However, one participant did suggest including references to the literature that informed each metric or category, stating:

"It would be nice to include in the metric framework the references to the literature which inspired each metric/category."

This comment reflects a desire for greater transparency and academic rigor, suggesting that while the metrics themselves are seen as appropriate, their foundational basis could be better documented. This would not only enhance the framework's credibility but also provide users with additional resources to understand the origins and justifications for each metric.

Another participant praised the systematic structure of the framework, which they felt effectively facilitated the measurement of diverse factors.

The comment stating:

"The metrics and categories in the framework facilitate the measurement of various diverse factors in a systematic structure."

This comment highlights the strength of the framework's organisation. This feedback suggests that the current organisation and categorisation are seen as logical and effective, contributing to the framework's overall usability.

5.4.4.3 Suggestions for Grouping and Labelling of Metrics

No suggestions were made regarding the grouping or labeling of metrics, metric categories, or subcategories, indicating that participants found the current structure to be satisfactory. One participant specifically commented:

"Grouping and labelling of metrics and categories is structured well and makes sense to me."

This reinforces the idea that the framework's organisational structure is intuitive and user-friendly, making it easy for users to navigate and apply the metrics in practice. The positive feedback on grouping and labeling suggests that the framework's design is accessible, reducing the likelihood of confusion or misinterpretation.

5.4.5 Overall Analysis of design validation

Overall, the feedback from participants indicates a high level of satisfaction with the TEDME-KG metrics framework. The absence of suggestions for new goals, metrics, or organisational changes suggests that the framework is seen as comprehensive, well-structured, and effective in its current form. The few suggestions made—such as adding references to the foundational literature—are focused on enhancing the transparency and academic rigor of the framework rather than altering its core components.

5.5 Evaluating general usability

In this section, the general usability of the metrics framework was assessed using the SUS score.

5.5.1 SUS score

The System Usability Scale (SUS) [135] is widely acknowledged as one of the most commonly utilised standardised questionnaires for evaluating the usability of a product or service following user testing. It consists of ten statements, as presented in Table 11, rated on a 5-point Likert scale, reflecting the extent of the respondent's agreement. These statements encompass several key aspects of usability, including the necessity for support, the system's complexity or simplicity, and the effectiveness of function integration. Due to its comprehensive coverage of these elements, the SUS is considered to possess high face validity in assessing the usability of a system.

The SUS is intentionally structured with a balance between positively and negatively worded items, where odd-numbered statements are phrased positively, and even-numbered statements are phrased negatively. Participants are instructed to respond to each item promptly, offering their initial reactions without excessive deliberation. It is essential that all ten items are answered; in cases where respondents are uncertain or unable to answer a particular item, they are advised to select the midpoint of the scale.

Table 16- System Usability Scale Questionnaire

Item	Item Description
1	I think that I would like to use this system frequently.
2	I found the system unnecessarily complex.
3	I thought the system was easy to use.
4	I think that I would need the support of a technical person to be able to use this system.
5	I found the various functions in this system were well integrated.
6	I thought there was too much inconsistency in this system.
7	I would imagine that most people would learn to use this system very quickly.
8	I found the system very cumbersome to use.
9	I felt very confident using the system.
10	I needed to learn a lot of things before I could get going with this system.

The SUS produces a single score that serves as a composite measure of the overall usability of the system being evaluated, without interpreting individual item scores separately. The final SUS score ranges from 0 to 100, with higher scores indicating better usability. Due to the alternating positive and negative phrasing of the statements, careful attention is required during the scoring process. Each item's score contribution, ranging from 0 to 4, must be calculated. For positively worded items (odd-numbered), the score contribution is derived by subtracting 1 from the scale position. For negatively worded items (even-numbered), the contribution is calculated as five minus the scale position. The sum of all score contributions is then multiplied by 2.5, resulting in an overall SUS score that ranges from 0 to 100 in increments of 2.5.

It is essential to clarify that, although the SUS score is frequently mistaken for a percentage, this interpretation is incorrect. Instead, SUS scores are more appropriately understood in terms of percentile ranges. With the accumulation of extensive datasets of SUS scores, several interpretation methods have been developed. A grading scale for SUS scores was first introduced, along with an adjective scale that uses descriptive terms rather than

numerical values. These scales were further refined through an extensive study, resulting in updated grading and adjective scales, as depicted in Figure 14.

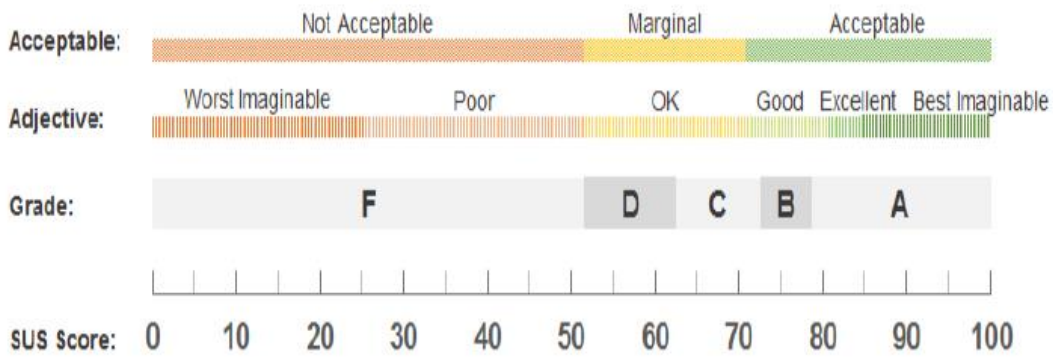


Figure 14- SUS score and its associated grading scales

Another approach to interpreting SUS scores involves categorizing them into three levels: "Acceptable," "Marginal," and "Not Acceptable." Scores above 71 are classified as "Acceptable," those ranging between 51.7 and 71 fall into the "Marginal" category, and scores below 51.7 are deemed "Not Acceptable." Furthermore, raw SUS scores can be converted into percentile ranks, as shown in studies that normalised large datasets of SUS scores, enabling comparisons. These percentile ranks indicate how a particular SUS score stands relative to others in the dataset.

The reliability of the SUS has been thoroughly validated. Early studies using coefficient alpha reported a reliability score of 0.85, and more recent research with larger sample sizes has demonstrated reliability exceeding 0.9.

Several key characteristics make the SUS a preferred choice among usability practitioners:

- **Technology-Agnostic:** The SUS can be used across a diverse range of technologies and services, offering great versatility.
- **Simplicity**: It is quick and straightforward to administer, making it convenient for both participants and researchers.
- **Non-Proprietary**: The SUS is publicly available and does not require any licensing fees for use.

These factors contribute to the widespread adoption and long-standing popularity of the SUS as a usability evaluation tool.

5.5.2 Procedure

A sample scenario was provided to the participants, where two temporal data modelling approaches were applied to a specific use case, resulting in the design of two KGs. Participants were asked to perform a series of tasks to evaluate the two approaches based on the framework. After completing the tasks, participants were asked to respond to 10 statements related to their experience and evaluation of the two approaches. Participants were asked to perform three tasks:

Task 1: Evaluate two different approaches based on the complexity in modelling category
Task 2: Evaluate two different approaches based on performance in querying
Task 3: Evaluate two different approaches based on the implementation and standardisation in uplifting

The original SUS items use the term "system," but various studies have demonstrated that replacing this with terms like "website," "product," or even the specific name of the product or website does not impact the scores, as long as the change is applied consistently across all items [136]. As noted by [137], making small adjustments to the wording in the SUS does not affect the results, provided the modifications remain uniform. For the purpose of this evaluation, all references to "system" were replaced with "Metrics framework" throughout the SUS questionnaire to better fit the context. The revised SUS questionnaire items used in this study are shown in Table 17.

Participants were then asked to indicate their agreement with each item on a five-point scale: (1) Strongly disagree; (2) Disagree; (3) Neither agree nor disagree; (4) Agree; (5) Strongly agree. Their responses were evaluated according to the specific SUS scoring process, as explained in detail in Section 5.4.5.1

Table 17- Amended version of the SUS questionnaire items used in this evaluation.

Item	Item Description
1	I think that I would like to use this the metrics framework frequently.
2	I found the metrics framework unnecessarily complex.
3	I thought the metrics framework was easy to use.

4	I think that I would need the support of a technical person to be able to use this metrics framework.
5	I found the various functions in this metrics framework were well integrated.
6	I thought there was too much inconsistency in this metrics framework.
7	I would imagine that most Semantic Web researchers would learn to use this metrics framework very quickly.
8	I found the metrics framework very cumbersome to use.
9	I felt very confident using the metrics framework.
10	I needed to learn a lot of things before I could get going with this metrics framework.

5.5.3 Results

Following the collection of responses from participants evaluating the TEDME-KG metrics framework, the System Usability Scale (SUS) scores were calculated. The mean SUS score across all participants was **85.5**. A python code was generated to calculate the SUS score (Appendix D)

5.5.3.1 Interpretation of the SUS Score

The mean SUS score of **85.5** was evaluated against established usability scales:

- **Grading Scale**: With a score of 85.5, the TEDME-KG metrics framework earns an "A" grade. This top-tier grade suggests that the framework is considered to have excellent usability by the participants.
- **Adjective Rating**: The SUS score places the framework in the "Excellent" to "Best Imaginable" category on the adjective scale. This indicates that users had a highly positive experience with the framework, finding it easy to use and highly effective in meeting their needs.
- **Acceptability**: The score of 85.5 significantly surpasses the threshold for acceptability (71), affirming that the TEDME-KG metrics framework is not only accepted by users but is also likely preferred over other systems they have encountered.

5.5.3.2 Descriptive Statistics and Confidence Interval

To provide a deeper understanding of the usability results, several descriptive statistics were calculated:

Table 18- Descriptive statistics results

Statistics	Value
Mean (μ)	85.5
Standard Deviation (σ)	10.33
Minimum	67.5
Maximum	100.0
Standard Error (SE)	3.27
Critical Value (t)	2.262
Margin of Error (ME)	7.39
95% Confidence Interval	78.11 - 92.89

The 95% confidence interval for the SUS score was calculated as follows:

Standard Error (SE): The standard error was calculated using the formula:

$$SE = \frac{10.33}{\sqrt{10}} \approx 3.27$$

Critical Value (t): For a 95% confidence level with 9 degrees of freedom (since n=10), the critical t-value was determined to be approximately 2.262.

Margin of Error (ME): The margin of error was calculated using the formula:

$$ME = t \times SE = 2.262 \times 3.27 \approx 7.39$$

Confidence Interval: The confidence interval was calculated as:

$$\text{Confidence Interval} = \bar{x} \pm ME = 85.5 \pm 7.39$$

This results in a confidence interval of 78.11 to 92.89**.**

This confidence interval suggests that we can be 95% confident that the true mean SUS score for the TEDME-KG metrics framework in the broader population lies between 78.11 and 92.89**.**

5.5.4 Overall Analysis of general usability

The analysis of the SUS scores for the TEDME-KG metrics framework reveals that it is perceived as highly usable. The framework's mean SUS score of 85.5 places it within the "Excellent" range on usability scales, and its "A" grade reflects outstanding user satisfaction. The confidence interval further supports this positive assessment, indicating that the true mean score for the population is likely to fall within a range that underscores the framework's exceptional usability.

These results suggest that the TEDME-KG metrics framework not only meets but exceeds user expectations, making it a robust and user-friendly tool in its intended application. The high usability score supports its potential for widespread adoption and success in real-world settings.

5.5.5 Limitations

The experimental evaluation of the proposed metrics framework involved only ten participants, which is a relatively small sample size. This limited number of responses restricts the statistical power and generalizability of the findings. Furthermore, three of the participants were PhD students with less than two years of hands-on experience in Semantic Web technologies, which may impact the depth and maturity of the feedback. Additionally, all participants were recruited through the researcher's supervisor network, raising concerns about potential selection bias. It is not explicitly known whether any respondents had a prior connection to the researcher, and this lack of transparency may influence the neutrality of the responses.

Despite these limitations, the study was designed to prioritize expert insight in a highly specialized domain. Expert-based evaluations in Semantic Web research often rely on small, focused samples where depth of knowledge is more critical than volume of responses. The inclusion of PhD students, while representing early-career experts, brought relevant and current perspectives from active research environments, complementing the views of more experienced participants. The recruitment method, while limited to the supervisor's network, is a common and pragmatic approach in academic research where domain-specific expertise is needed. To reduce potential bias, the survey was administered anonymously, allowing participants to respond freely and without influence. These measures were taken to ensure

that, even within a constrained sample, the feedback remained credible, diverse, and relevant to the aims of the study.

5.6 Chapter Summary

In this chapter, the TEDME-KG metrics framework was evaluated based on feedback from experts in the Semantic Web field. The evaluation focused on three main areas: the appropriateness of the framework for evaluating temporal data modelling, the general usability of the framework, and the experts' suggestions for improvement.

An anonymous survey was conducted with 10 expert participants, consisting of six PhD researchers and four postdoctoral researchers. These experts provided feedback on the framework through three sections. The first section assessed the appropriateness of the framework using criteria such as relevance, clarity, directness, feasibility, understandability, and comprehensiveness. Overall, the feedback was highly positive. Experts found the framework to be relevant, clear, and operationally feasible, with a few suggestions for improving clarity in certain areas.

The evaluation also explored suggestions for new goals, metrics, and the grouping of metrics. However, no major changes were recommended, indicating that the experts found the framework comprehensive and well-structured. One suggestion included referencing the literature to enhance the transparency and academic rigor of the framework.

Additionally, the general usability of the framework was assessed using the System Usability Scale (SUS). The mean SUS score was 85.5, placing the framework in the "Excellent" to "Best Imaginable" range, earning it an "A" grade in usability. This high score indicates that the framework is not only functional but also easy to use, aligning well with user expectations.

Overall, the chapter concludes that the TEDME-KG metrics framework is both appropriate and highly usable, with minor improvements suggested for clarity and transparency. The positive feedback highlights its potential for successful adoption in real-world applications.

Chapter 6 Implementation

In this chapter, a comprehensive experimental study was conducted, focusing on the evaluation of various temporal data modelling approaches. The study was centred around the FAIRVASC use case, which provided a real-world context. These approaches were systematically evaluated using the TEDME-KG metrics framework.

6.1 Methodology

The primary hypothesis for this study is:

“The TEDME-KG metrics framework can evaluate various temporal data modelling approaches in knowledge graphs based on modelling, uplifting and querying aspects.”

This hypothesis stems from the goal of evaluating different RDF-based temporal modelling approaches through a holistic metrics framework. The framework will assess three key categories: Modelling, Uplifting, and Querying, which together address the central research question of this thesis: *To what extent can a metrics framework facilitate the evaluation of temporal data modelling approaches in knowledge graphs for Semantic Web researchers?*

Six phases and their associated artefacts were proposed to evaluate temporal data modelling approaches in KGs. Each phase is designed to contribute to the evaluation of temporal data modelling approaches in KGs. Figure 15 depicts the phases. The evaluation phases are displayed in black, evaluation actions are displayed in white rectangles, while the artifacts are highlighted within grey shapes.

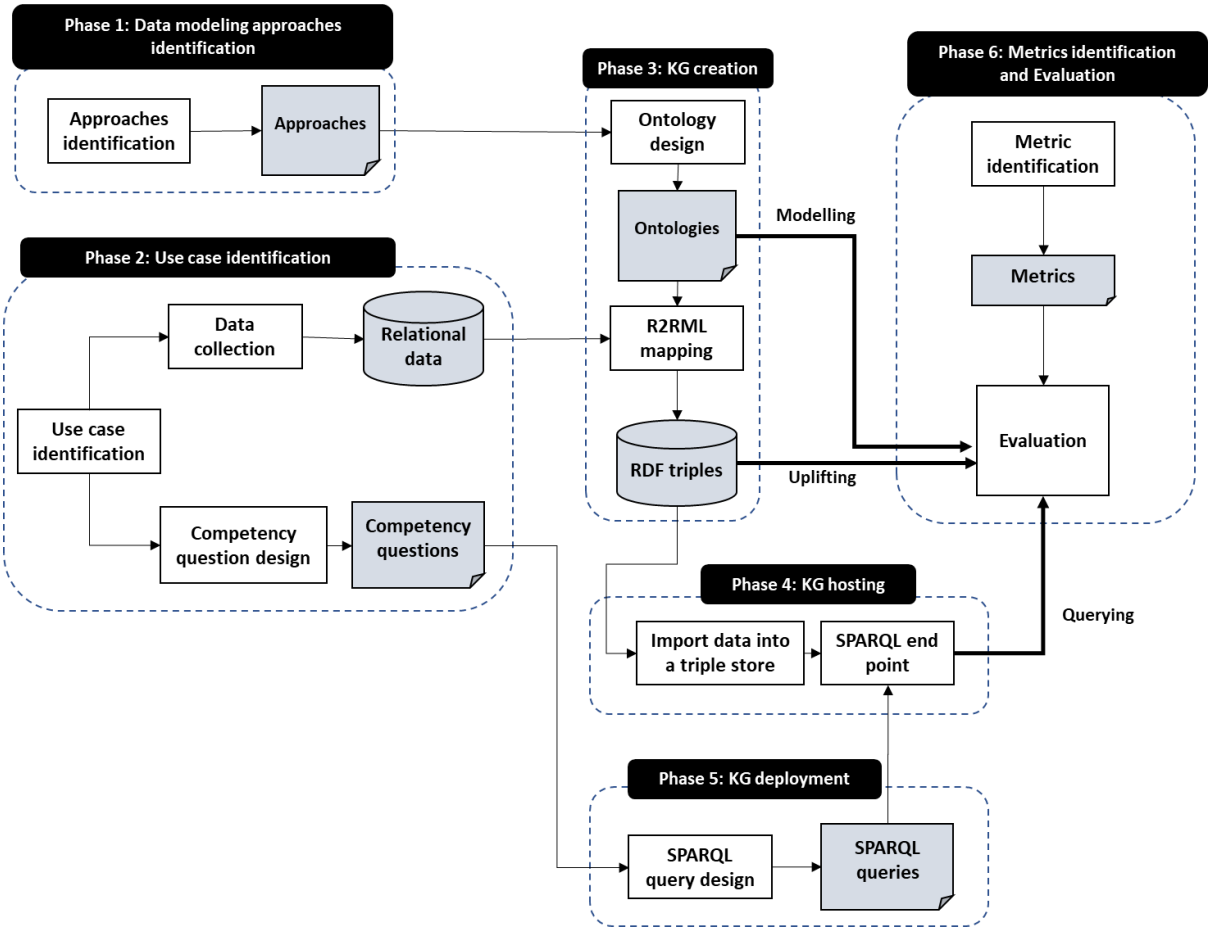


Figure 15- The phases for evaluating temporal data modelling approaches in KGs

The phases and their associated practical implementation details are discussed below.

6.1.1 Phase 1: Data Modelling Approaches Identification

The first phase of this process involved identifying various temporal data modelling approaches in RDF-based knowledge graphs (KGs). In Chapter 3, a systematic review was conducted, leading to the identification of several temporal data modelling approaches. Building on the findings from Chapter 3, seven of these approaches were selected for detailed evaluation in this experimental study.

6.1.2 Phase 2: Use Case Identification

In this phase, the specific use case is defined by collecting relevant data and designing competency questions. The use case was discussed in chapter 2. The competency questions guide the structure and objectives of the knowledge graph, ensuring it meets the needs of the use case. The competency questions are discussed below.

To cover variety of temporal question types, a question was considered for each category and each time entity. (Table 19)

Table 19- Temporal question categories

Temporal question type	Time entities	Temporal competency question
1. Explicit temporal	Time point	Q1: Which patients used drug with ncit:C769 as a treatment on 2019-08-01?
	Time interval	Q2: Which patients used drug with ncit:C769 as a treatment between 2016-07-18 until 2016-07-25?
2. Implicit temporal	Time Interval	Q3: Which medications used by patient=12, 1 month before and 1 month after his/her relapse?
	Time point	Q4: Which medications used by patient=12 before his/her relapse?
3. Ordinal temporal	Time interval	Q5: What is the duration of second use of drug with ncit:C769 for patient with ID=12?
	Time point	Q6: What is the date of second use of drug with ncit:C769 for patient with RKDID=40?
4. Temporal answers	Time interval	Q7: What is the duration of using a drug with ncit:C769 for a patient with ID=12?
	Time point	Q8: What are the dates of use of drug with ncit:C769 for patient with ID=12?

6.1.3 Phase 3: KG Creation

To bridge theory and application, a temporal ontology was designed based on identified approaches, RDF data was generated using the R2RML engine using the designed ontology. This phase formed the backbone for subsequent KG hosting and deployment.

6.1.4 Phase 4: KG Hosting

The RDF triples were imported into a triple store, and a SPARQL endpoint was set up.

6.1.5 Phase 5: KG deployment

With the RDF data hosted, SPARQL queries were designed aligning with the competency questions. These queries form the basis for evaluating the performance of RDF-based KG modelling approaches.

6.1.6 Phase 6: Metrics Identification and Evaluation

In the final phase, based on the TEDME-KG metrics framework the temporal data modelling approaches were evaluated.

6.2 Technical Details

This section provides a comprehensive technical detail of the implementation process for the selected approaches, following the evaluation phases outlined in Section 6.1.

Based on the results of the systematic review presented in chapter 3, six temporal data modelling approaches were chosen. For each approach, a KG was designed and evaluated using the TEDME-KG metrics framework.

In total, seven ontologies were developed, seven R2RML mapping codes were created, and 56 SPARQL queries were designed to address various temporal question categories, with eight queries per approach.

This section presents an ontology, a snippet of the uplift code, and the SPARQL queries for a few approaches as an example. All SPARQL queries have been uploaded to a GitHub repository. https://github.com/sepidha/PhD/tree/Thesis

6.2.1 Experiment setup

For the experiment's setup, the data in CSV format were uplifted using the R2RML-F engine. The generated RDF datasets were then loaded onto a GraphDB triple store running on a single PC with

Based on the six phases discussed in Section 6.1 and following the use cases and approaches covered in previous chapters, we will now focus on phases 3, 4, 5, and 6.

6.2.2 Ontologies

6.2.2.1 Approach 1: Named Graph

An ontology was designed based on the use case and the Named Graph approach. In this approach, we consider all the elements inside the red dashed boundary as a part of a single

named graph. The start date and stop date (dateTime) can then be attached to this named graph, representing the temporal context for the entire graph.

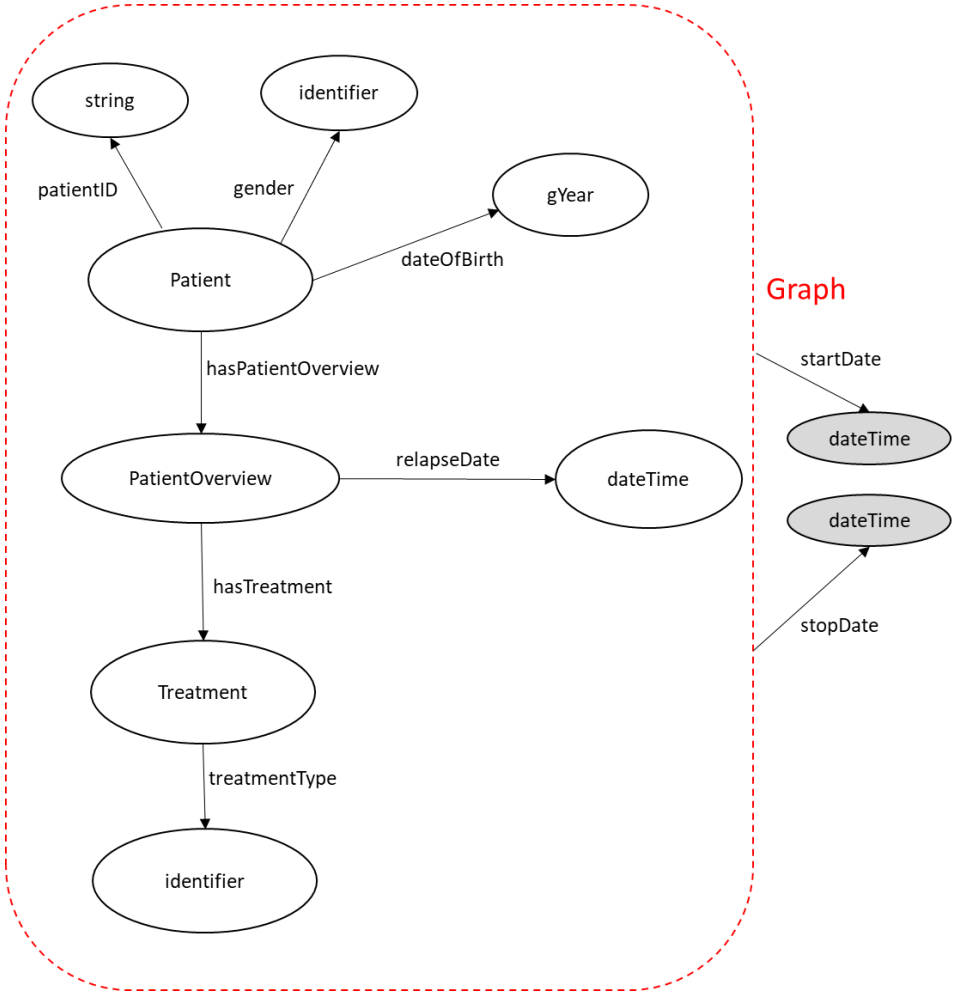


Figure 16- Adding the medication start and stop date to the Named Graph approach

6.2.2.2 Approach 2: Standard Reification

An ontology was designed based on the use case and the standard Reification approach. The reification pattern is applied by introducing a statement node that reifies the relationship between a patient and their treatment, allowing additional temporal information (startDate and stopDate) to be attached to these relationships.

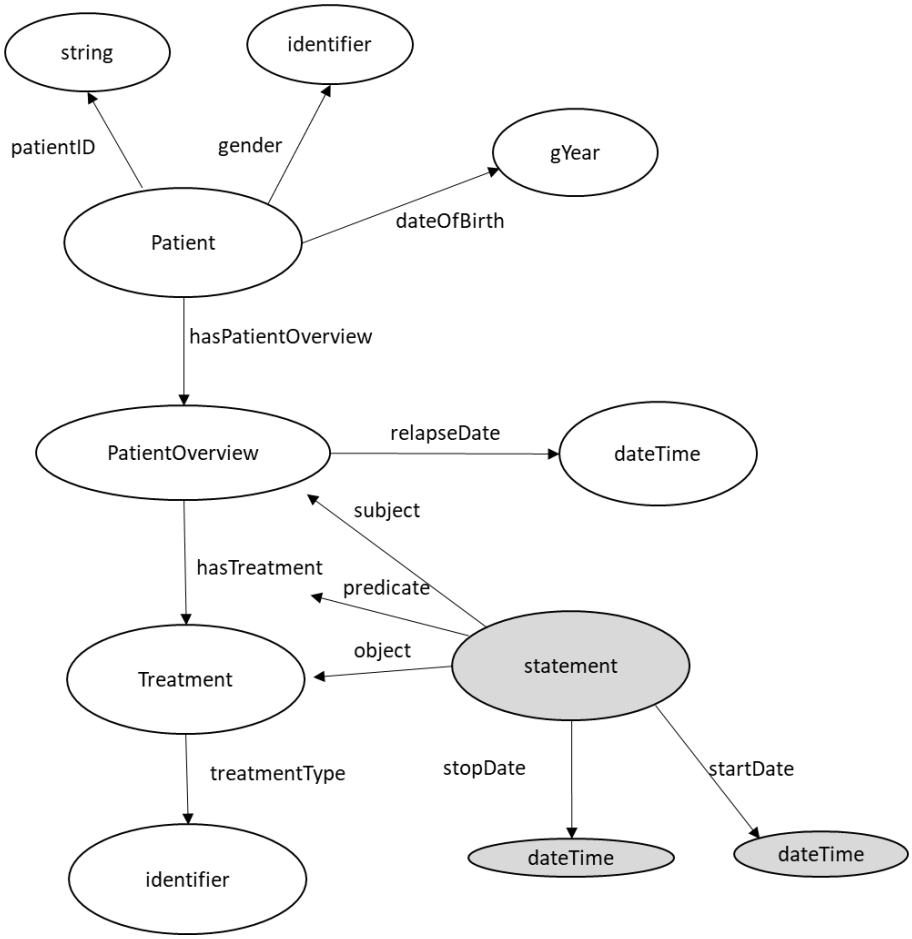


Figure 17- Adding the medication start and stop date to the Standard Reification approach.

6.2.2.3 Approach 3: RDF-Star

An ontology was designed based on the use case and the RDF-Star approach. In this approach, we consider a triple inside the red dashed boundary as a subject of the second triple. The start date and stop date can then be attached as predicates and dateTime as objects in the second triple, representing the temporal context for the triple.

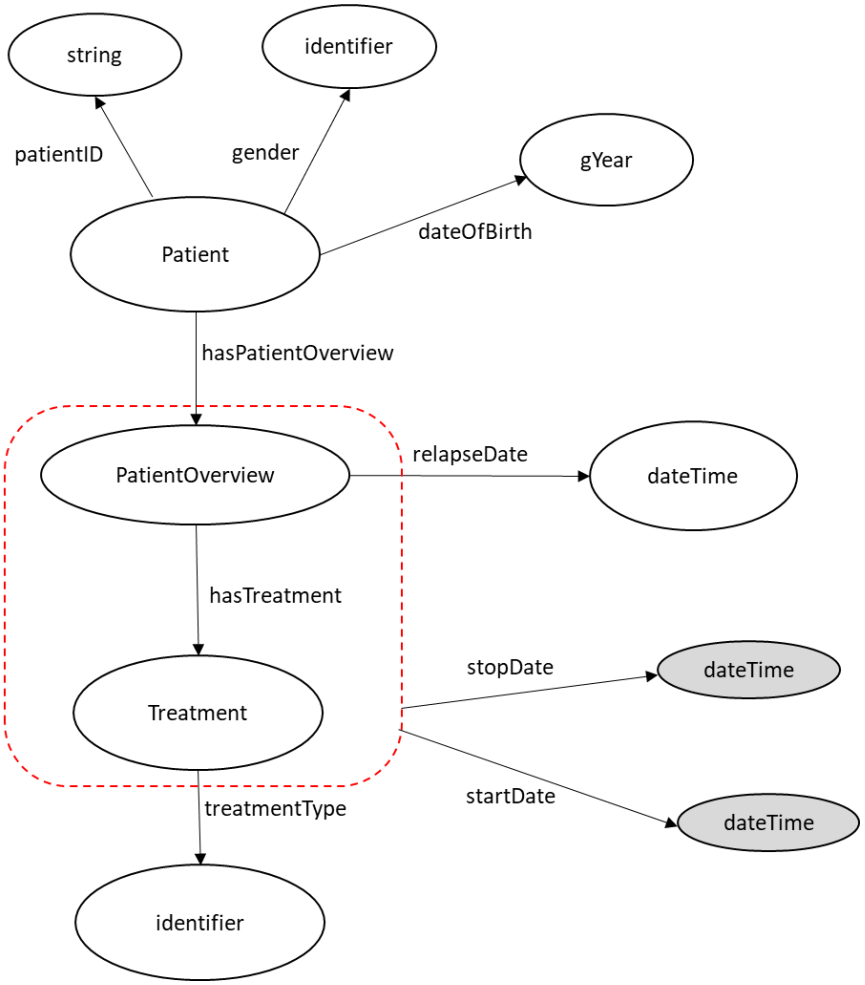


Figure 18- Adding the medication start and stop date to the RDF-Star approach.

6.2.2.4 Approach 4: Temporal RDF Graph

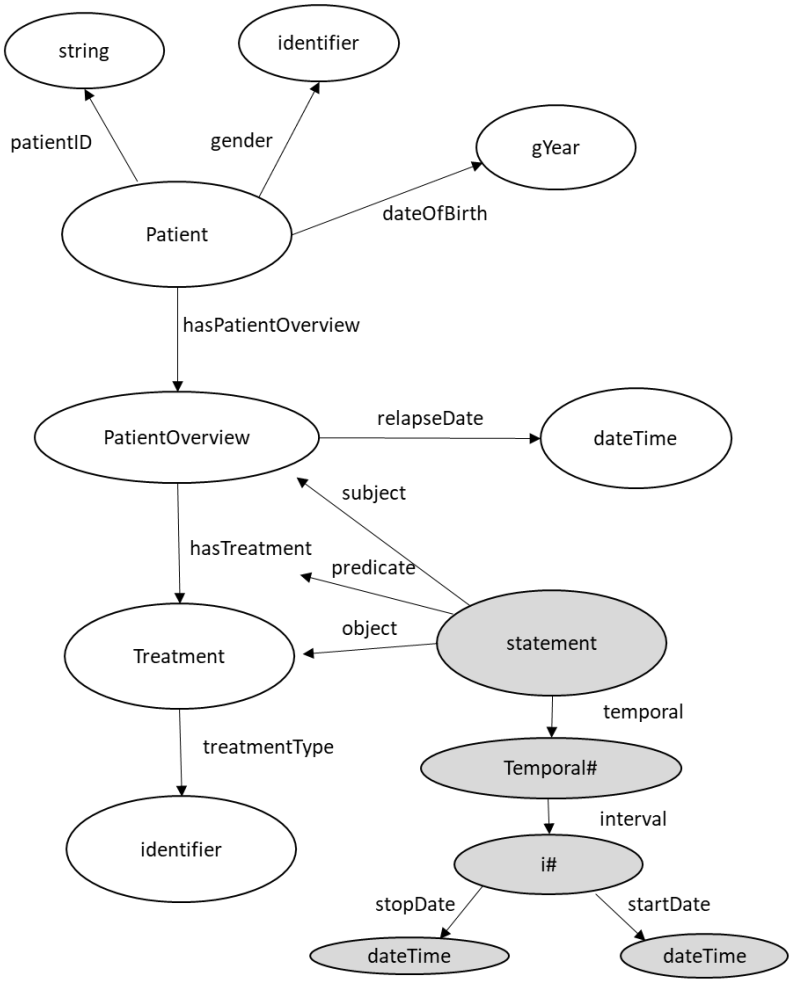


Figure 19- Adding the medication start and stop date to the Temporal RDF Graph approach

6.2.2.5 Approach 5: RDFMT

An ontology was designed based on the use case and the RDFMT approach. In this approach a node called statement was created and linked to a triple, which represented the relationship between a Patient and a Treatment (through hasTreatment). The statement node serves as a reified version of the triple, encapsulating the subject, predicate, and object in a way that allows metadata to be attached to this specific triple. For each part of the triple (subject, predicate, and object), additional metadata like start date and end date were attached.

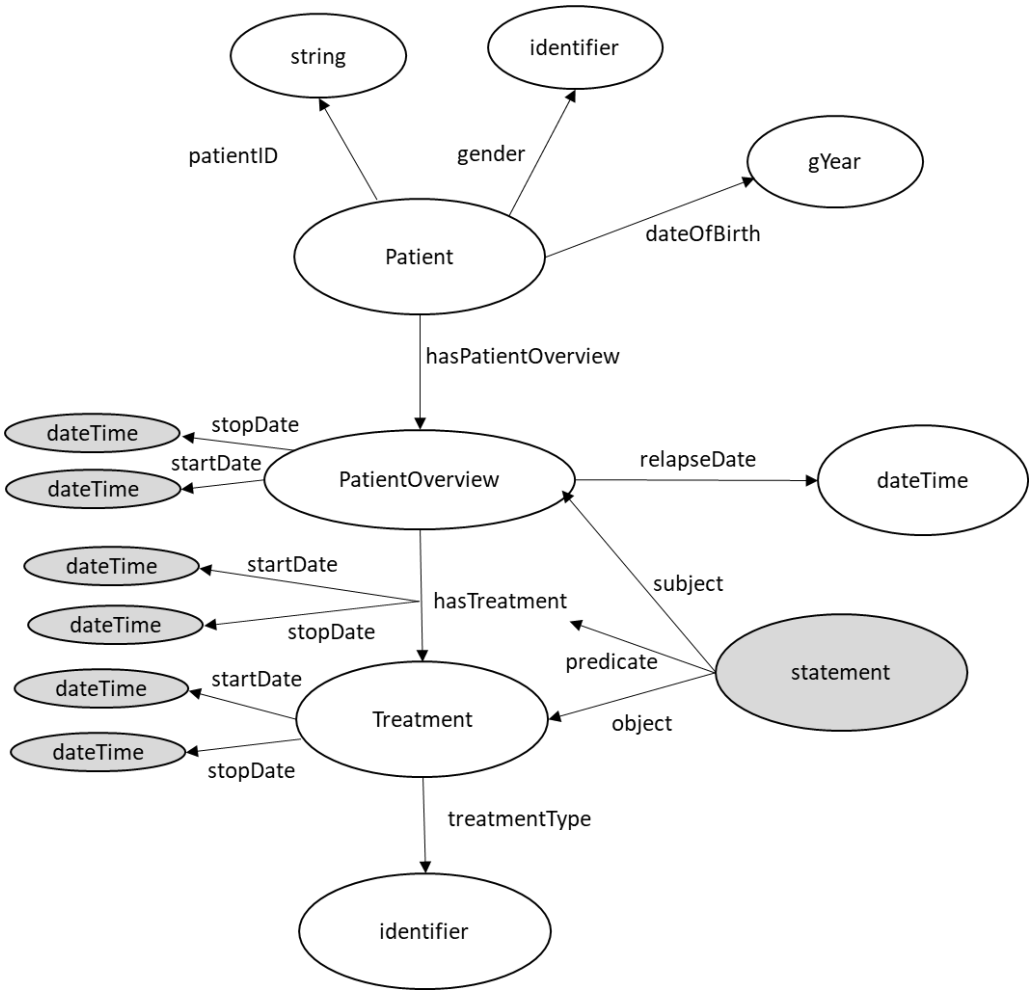


Figure 20- Adding the medication start and stop date to the RDFMT approach

6.2.2.6 Approach 6: RDFt

An ontology was designed based on the use case and the RDFt approach. In this approach, a statement node is introduced to represent (PatinetOverview, has treatment, treatment). The stop date and start date of the treatment is attached to the predicate (hasTreatment).

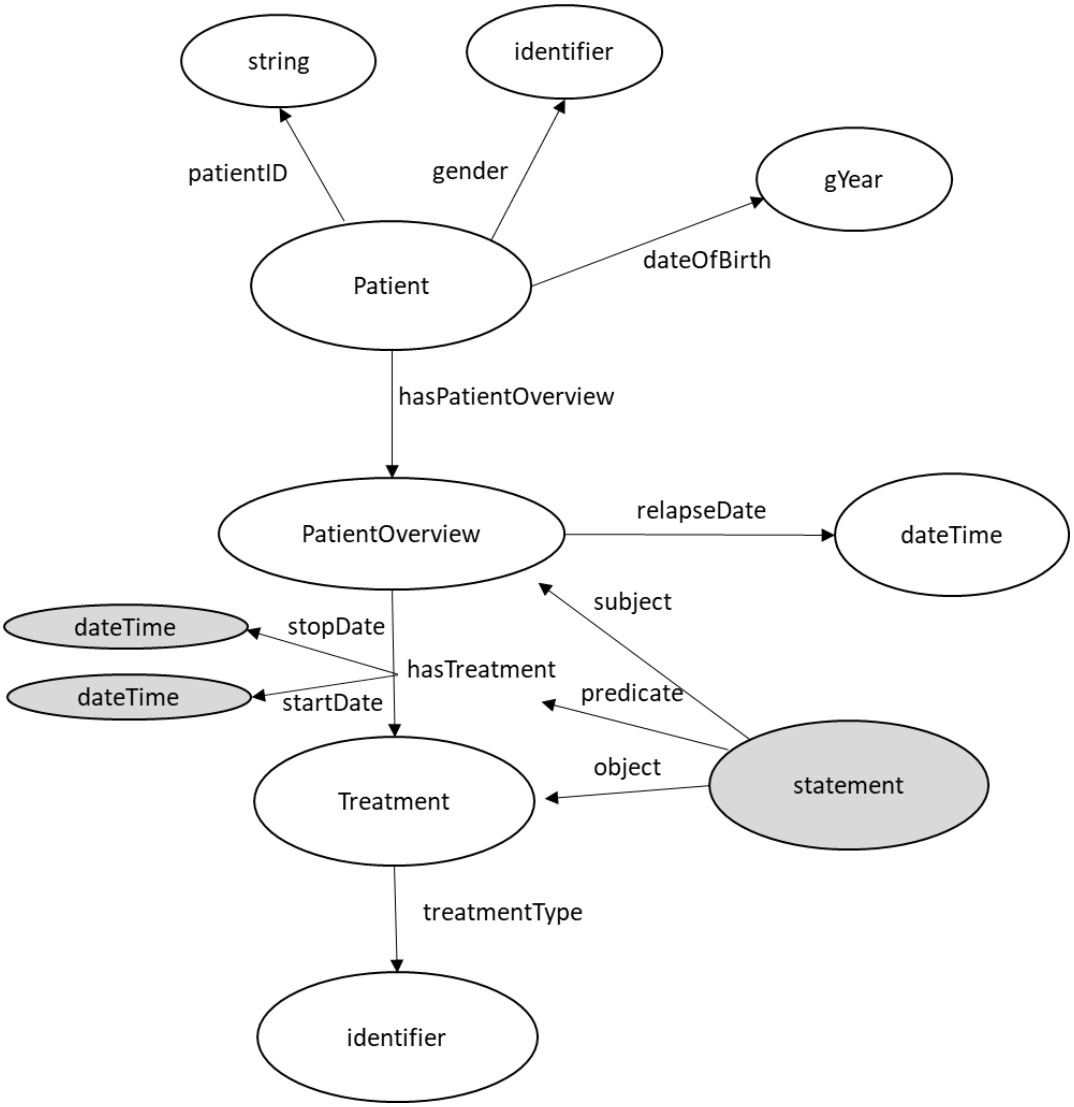


Figure 21- Adding the medication start and stop date to the RDFt approach.

6.2.2.7 Approach 7: Singleton Property

An ontology was designed based on the use case and the singleton property approach. In this approach, a unique predicate of hasTreatment was created to represent a specific occurrence of a treatment for a patient. This unique predicate is a singleton property of the general hasTreatment relationship. additional metadata, such as the start date and stop date were then added to the unique hasTreatment.

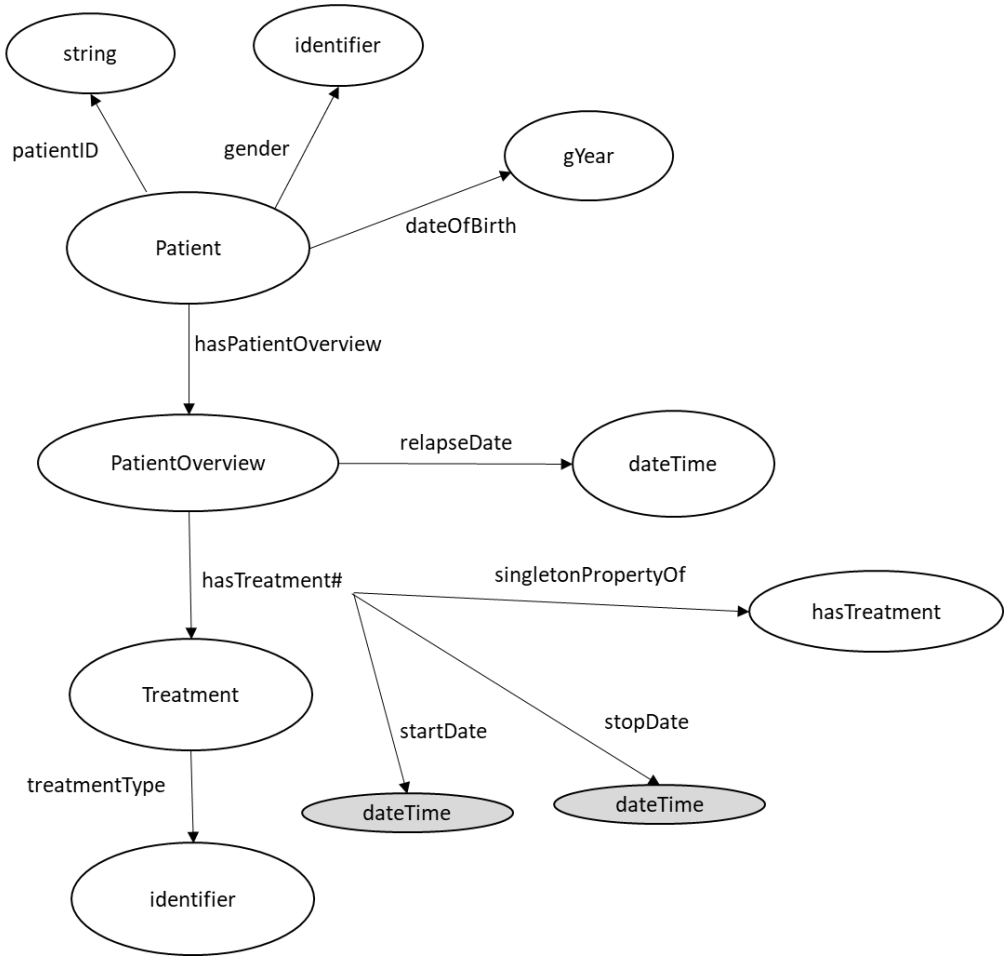


Figure 22- Adding the medication start and stop date to the Singleton Property approach

6.2.3 R2RML Mappings

Once the ontology was created, the data was uplifted based on the ontology using the R2RML engine. A snippet of the R2RML code for the Named Graph as an example is provided below:

Listing 1- Snippet of the R2RML code based on the Named Graph approach

```
<#patient_overview>

  rr:logicalTable [ rr:sqlQuery """SELECT
    RECORD_ID,
    STOP_DATE||'T00:00:00' AS STOP_DATE,
    START_DATE||'T00:00:00' AS START_DATE,
    DRUG_SELECTED_CHOICES,
    RELAPSE_DATE ||'T00:00:00' AS RELAPSE_DATE
  FROM RKD""" ];
```

```
  rr:subjectMap [
    rr:template
"http://w3id.org/FAIRVASC/resource/rkd/patient/patientoverview/{RECORD_ID}
/{START_DATE}{STOP_DATE}" ;
    rr:class fvc:PatientOverview;
    ## named graph
    rr:graph <http://w3id.org/FAIRVASC/graph> ;
  ] ;
```

The core of the mapping is the rr:logicalTable, which defines an SQL query to select relevant data from the RKD table, including the record ID, start and stop dates of treatments, and a DRUG_ENCODING value derived from a CASE statement that maps drug choices to specific codes. The dates are formatted as date-time strings to ensure proper temporal representation in RDF. The rr:subjectMap generates unique URIs for each RDF resource by combining the record ID, drug encoding, start date, and stop date into a single URI template, indicating that each resource represents a specific treatment event. This resource is then assigned the class fvc:statement. Two rr:predicateObjectMap elements map the START_DATE and STOP_DATE columns to the RDF properties fvc:StartDate and fvc:StopDate, respectively, ensuring these dates are correctly typed as xsd:dateTime in the resulting RDF data.

6.2.4 SPARQL Queries

Based on the competency questions discussed in Section 5.8, eight SPARQL queries were designed. For each approach, 8 Sparql queries (56 queries in total) were created to test a range of temporal queries.

In this section, for each temporal question category, a snippet of the code is provided based on one or two approaches. The remaining queries can be found in the GitHub repository. https://github.com/sepidha/PhD/tree/Thesis

6.2.4.1 Explicit temporal (Time Point) - EXTP

The question in this category is as follows:

Q1: Which patients used prednisone as a continuing medication treatment on 2019-08-01?

The Sparql query for the Named Graph approach as an example is presented below.

Listing 2- Sparql query for the Named Graph approach for the explicit temporal (time point) category

```
#---------------------------- Namespaces----------------------------------
--

prefix fvc: <http://w3id.org/FAIRVASC#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix xsd: <http://www.w3.org/2001/XMLSchema#>
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
Select distinct ?patientID ?StartDate ?StopDate
from Named <http://w3id.org/FAIRVASC/graph>
Where {
  GRAPH <http://w3id.org/FAIRVASC/graph> {
?patient a fvc:Patient;
fvc:patientID "4";
fvc:hasPatientOverview ?patientoverview.
?patientoverview fvc:hasTreatment ?treatment.
?treatment fvc:treatmentType <http://identifiers.org/ncit:C769>.
?patientoverview fvc:StartDate ?StartDate.
?patientoverview fvc:StopDate ?StopDate.
FILTER (?StartDate >= "2017-08-01T00:00:00"^^xsd:dateTime) }}
```

Listing 3- Sparql query for the RDF—Star approach for the explicit temporal (time point) category

```
PREFIX fvc: <http://w3id.org/FAIRVASC#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02-rdf-syntax-ns#>

SELECT DISTINCT ?patient
WHERE {
  ?patient a fvc:Patient;
           fvc:patientID ?patientID;
           fvc:hasPatientOverview ?patientoverview.

  << ?patientoverview fvc:hasTreatment ?treatment >>
    fvc:StartDate ?StartDate;
    fvc:StopDate ?StopDate.

  ?treatment fvc:treatmentType <http://identifiers.org/ncit:C769>.

  FILTER ("2019-08-01"^^xsd:date >= ?StartDate && "2019-08-01"^^xsd:date
<= ?StopDate)
}
```

6.2.4.2 Explicit temporal (Time Interval) - EXTI

```
PREFIX fvc: <http://w3id.org/FAIRVASC#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT DISTINCT  ?treatmentType ?relapseDate ?StartDate ?StopDate
WHERE {?patient a fvc:Patient;
fvc:patientID "12";
fvc:hasPatientOverview ?patientoverview.
?patientoverview fvc:relapseDate ?relapseDate.
?patientoverview fvc:hasTreatment ?treatment.
?treatment fvc:treatmentType ?treatmentType.
?statement rdf:subject ?patientoverview.
?statement rdf:predicate ?hasTreatment.
?statement rdf:object ?treatment.
?statement fvc:StartDate ?StartDate.
?statement fvc:StopDate ?StopDate.
FILTER (?StopDate <= ?relapseDate)}
```

The question in this category is as follows:

Which patients used prednisone as a continuing medication between 2016-07-18 until 2016-07-25?

Since this query refers to a time interval, it functions similarly to a time point query, but the filtering options adjust to account for the time range instead of a single point in time.

Listing 4- Sparql query for the Named Graph approach for the explicit temporal (time interval) category

```
FILTER ((?StartDate >= "2019-08-01T00:00:00"^^xsd:dateTime) && (?StopDate
<= "2020-08-01T00:00:00"^^xsd:dateTime))}
```

6.2.4.3 Implicit temporal (Time Point) - IMTP

The question in this category is as follows:

Which medications used by patient=12 before his/her relapse?

The Sparql query for the Standard Reification approach as an example is presented below.

Listing 5- Sparql query for the Standard Reification approach for the Implicit temporal (time point) category

6.2.4.4 Implicit Temporal (Time Interval) - IMTI

The question in this category is as follows:

Which medications used by patient=12, 1 month before and 1 month after his/her relapse?

The Sparql query for the Standard Reification approach as an example is presented below.

Listing 6- Sparql query for the Standard Reification approach for the Implicit temporal (time interval) category

```
PREFIX fvc: <http://w3id.org/FAIRVASC#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT DISTINCT  ?treatmentType ?relapseDate ?StartDate ?StopDate
WHERE {  ?patient a fvc:Patient;
fvc:patientID "12";
fvc:hasPatientOverview ?patientoverview.
?patientoverview fvc:relapseDate ?relapseDate.
BIND(xsd:dateTime(?relapseDate) - "P1M"^^xsd:duration AS ?oneMonthBefore)
BIND(xsd:dateTime(?relapseDate) + "P1M"^^xsd:duration AS ?oneMonthAfter)
?patientoverview fvc:hasTreatment ?treatment.
?treatment fvc:treatmentType ?treatmentType.
?statement rdf:subject ?patientoverview.
?statement rdf:predicate ?hasTreatment.
?statement rdf:object ?treatment.
?statement fvc:StartDate ?StartDate.
?statement fvc:StopDate ?StopDate.
FILTER (?StartDate >= ?oneMonthBefore && ?StopDate <= ?oneMonthAfter)}
```

Listing 7- Sparql query for the RDF-Star approach for the Implicit temporal (time interval) category

```
PREFIX fvc: <http://w3id.org/FAIRVASC#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT DISTINCT ?treatmentType ?relapseDate ?StartDate ?StopDate
WHERE {
  ?patient a fvc:Patient;
           fvc:patientID "12";
           fvc:hasPatientOverview ?patientoverview.
  ?patientoverview fvc:relapseDate ?relapseDate.

  # Calculate 1 month before and after
  BIND(xsd:date(?relapseDate) - "P1M"^^xsd:duration AS ?oneMonthBefore)
  BIND(xsd:date(?relapseDate) + "P1M"^^xsd:duration AS ?oneMonthAfter)
```

```
  << ?patientoverview fvc:hasTreatment ?treatment >>
    fvc:StartDate ?StartDate;
    fvc:StopDate ?StopDate.
  ?treatment fvc:treatmentType ?treatmentType.

  FILTER (?StartDate >= ?oneMonthBefore && ?StopDate <= ?oneMonthAfter)
}
```

6.2.4.5 Ordinal Temporal (Time Point) - ORTP

The question in this category is as follows:

What is the date of second use of drug with ncit:C769 for patient with RKDID=40?

The Sparql query for the Singleton Property approach as an example is presented below.

Listing 8- Sparql query for the Singleton Property approach for the ordinal temporal (time point) category

```
prefix fvc: <http://w3id.org/FAIRVASC#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix xsd: <http://www.w3.org/2001/XMLSchema#>
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ofn: <http://www.ontotext.com/sparql/functions/>
Select distinct ?patientID ?relapseDate ?StartDate ?StopDate
Where { ?patient a fvc:Patient;
fvc:patientID ?patientID;
fvc:hasPatientOverview ?patientoverview.
    # Retrieve the relapse event and date
  ?PatientOverview fvc:relapseDate ?relapseDate.
?patientoverview fvc:hasTreatment ?treatment.
?treatment fvc:treatmentType <http://identifiers.org/ncit:C769>.
?hasTreatment rdf:singletonPropertyOf ?hasTreatments.
?hasTreatment fvc:StartDate ?StartDate.
?hasTreatment fvc:StopDate ?StopDate.
# Ensure the medication use starts after the relapse date
  FILTER (?StartDate > ?relapseDate)}
ORDER BY ?StartDate
LIMIT 1   # Limit to the first result
```

Listing 9- Sparql query for the RDF-Star approach for the ordinal temporal (time point) category

```
PREFIX fvc: <http://w3id.org/FAIRVASC#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02-rdf-syntax-ns#>

SELECT DISTINCT ?treatmentType ?relapseDate ?StartDate
WHERE {
  ?patient a fvc:Patient;
           fvc:patientID "12";
           fvc:hasPatientOverview ?patientoverview.
  ?patientoverview fvc:relapseDate ?relapseDate.

  << ?patientoverview fvc:hasTreatment ?treatment >>
    fvc:StartDate ?StartDate.
  ?treatment fvc:treatmentType ?treatmentType.

  FILTER (?StartDate > ?relapseDate)
}
ORDER BY ?StartDate
LIMIT 1
```

6.2.4.6 Ordinal Temporal (Time Interval) - ORTI

The question in this category is as follows:

What is the duration of second use of drug with ncit:C769 for patient with ID=12?

The Sparql query for the Singleton Property approach as an example is presented below.

Listing 10- Sparql query for the Singleton Property approach for the ordinal temporal (time interval) category

```
prefix fvc: <http://w3id.org/FAIRVASC#>
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix xsd: <http://www.w3.org/2001/XMLSchema#>
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ofn: <http://www.ontotext.com/sparql/functions/>
Select distinct ?patientID ?StartDate ?StopDate ?DurationDays
Where { ?patient a fvc:Patient;
fvc:patientID ?patientID;
fvc:hasPatientOverview ?patientoverview.
    # Retrieve the relapse event and date
  ?PatientOverview fvc:relapseDate ?relapseDate.
?patientoverview fvc:hasTreatment ?treatment.
?treatment fvc:treatmentType <http://identifiers.org/ncit:C769>.
?hasTreatment rdf:singletonPropertyOf ?hasTreatments.
?hasTreatment fvc:StartDate ?StartDate.
?hasTreatment fvc:StopDate ?StopDate.
 # Calculate the duration in days between StartDate and StopDate
  BIND(ofn:daysBetween(?StartDate, ?StopDate) AS ?DurationDays)}
ORDER BY ?StartDate
OFFSET 1  # Skip the first instance
LIMIT 1   # Limit to the second instance
```

6.2.4.7 Temporal Answer (Time point) - TETP

The question in this category is as follows:

What are the dates of use of drug with ncit:C769 for patient with ID=12?

The Sparql query for the Temporal RDF Graph approach as an example is presented below.

Listing 11- Sparql query for the Temporal RDF Graph approach for the temporal answer (time point) category

```
PREFIX fvc: <http://w3id.org/FAIRVASC#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ofn: <http://www.ontotext.com/sparql/functions/>
SELECT DISTINCT ?StartDate ?StopDate
WHERE {
  # Identify the patient with patientID 12
  ?patient a fvc:Patient;
           fvc:patientID "12";
           fvc:hasPatientOverview ?PatientOverview.
  # Retrieve treatments for the specific drug ncit:C769
```

```
  ?PatientOverview fvc:hasTreatment ?treatment.
  ?treatment fvc:treatmentType <http://identifiers.org/ncit:C769>.
   ?statement rdf:subject ?PatientOverview.
  ?statement fvc:hastemporal ?temporal.
  ?temporal fvc:hasinterval ?i.
  ?i fvc:StartDate ?StartDate;
     fvc:StopDate ?StopDate. }
```

6.2.4.8 Temporal Answer (Time Interval) – TETI

The question in this category is as follows:

What is the duration of using a drug with ncit:C769 for a patient with ID=12?

The Sparql query for the Temporal RDF Graph approach as an example is presented below.

Listing 12- Sparql query for the Temporal RDF Graph approach for the temporal answer (time interval) category

```
PREFIX fvc: <http://w3id.org/FAIRVASC#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX ofn: <http://www.ontotext.com/sparql/functions/>
SELECT DISTINCT ?StartDate ?StopDate ?DurationDays
WHERE {  ?patient a fvc:Patient;
           fvc:patientID "12";
           fvc:hasPatientOverview ?PatientOverview.
  ?PatientOverview fvc:hasTreatment ?treatment.
  ?treatment fvc:treatmentType <http://identifiers.org/ncit:C769>.
   ?statement rdf:subject ?PatientOverview.
  ?statement fvc:hastemporal ?temporal.
  ?temporal fvc:hasinterval ?i.
  ?i fvc:StartDate ?StartDate;
     fvc:StopDate ?StopDate.
  BIND(ofn:daysBetween(?StartDate, ?StopDate) AS ?DurationDays)}
ORDER BY ?StartDate
```

Listing 13- Sparql query for the RDF-Star approach for temporal answer (time interval) category

```
PREFIX fvc: <http://w3id.org/FAIRVASC#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02-rdf-syntax-ns#>
PREFIX ofn: <http://www.ontotext.com/sparql/functions/>
```

```
SELECT DISTINCT ?Duration
WHERE {
  ?patient a fvc:Patient;
           fvc:patientID "12";
           fvc:hasPatientOverview ?patientoverview.

  << ?patientoverview fvc:hasTreatment ?treatment >>
    fvc:StartDate ?StartDate;
    fvc:StopDate ?StopDate.
  ?treatment fvc:treatmentType <http://identifiers.org/ncit:C769>.

  # Calculate duration
  BIND((xsd:date(?StopDate) - xsd:date(?StartDate)) AS ?Duration)
}
```

6.2.5 Discussion on SPARQL queries

While preparing SPARQL queries for all question categories, the same set of questions was used across all approaches to ensure comparability. Redundant lines were removed from the queries, and SPARQL functions were applied consistently across all models being tested. A systematic approach was adopted to vary query complexity by designing different categories of questions that reflect typical temporal reasoning tasks, such as calculating durations, comparing timestamps (e.g., before/after, first/second events), and using built-in SPARQL functions like BIND, FILTER, and arithmetic expressions. This method allowed for controlled increases in logical and structural query complexity.

Query length was measured by counting the number of distinct lines in the SPARQL query, where each line was terminated by a period (.), indicating the end of a triple pattern or directive. This method provides a consistent way to assess the structural complexity of the queries, independent of whitespace or formatting variations.

6.3 Evaluation results

The approaches were evaluated using the TEDME-KG metrics framework. For easier readability, abbreviations of the approaches were used in the evaluation results:

Table 20- Acronyms for each temporal data modelling approaches

Approach	Acronym
Standard Reification	SR
Named Graph	NG

RDFt	Rt
RDFMT	RM
Singleton Property	SP
RDF-Star	RS
Temporal RDF graph	TRG

6.3.1 Modelling metrics category

Firstly, the modelling metrics category was considered. This category includes two key subcategories: implementation support and standardisation, and complexity.

6.3.1.1 Implementation support and standardisation

Implementation support and standardisation were assessed using two metrics: ontology editor support and adherence to W3C standards. Protégé, a widely used ontology editor, was tested to model the ontologies, and W3C recommendations were reviewed to determine if the approaches adhered to these standards. The results are presented in the table below:

Table 21- Implementation support and standardisation

Metrics category	Metrics	1. SR	2. RM	3. TG	4. Rt	5. SP	6. RS	7. NG
Implementation support and standardisation	Ontology editor support	No	No	No	No	No	No	No
	Adherence to W3C	No	No	No	No	No	Yes	Yes

The results for implementation support and standardization indicate that none of the approaches currently offer support within ontology editors, which poses challenges for practical adoption. However, only the RS (RDF-star) and NG (Named Graph) approaches—represented as (7) and (8), respectively—align with W3C standards or are on a clear path toward standardization. NG is already part of the official W3C specifications, while RS is under active development within the W3C RDF-star Working Group and is progressing toward becoming a formal recommendation. This alignment is crucial for ensuring interoperability and widespread acceptance within Semantic Web technologies.

6.3.1.2 Complexity

Complexity was calculated by running a query in Graphdb. (Listing 14) The first part of the query counts the total number of triples (?Triples) by selecting all subject-predicate-object combinations in the dataset. It also counts the total number of distinct nodes (?TotalNodes), which are entities that appear either as subjects or objects in any triple. This is done by using a UNION to ensure all unique nodes are considered, regardless of their role in the triples.

The query then uses these counts to calculate redundancy and additional triples by comparing the current dataset to a baseline. The ?Redundancy value is calculated by adding the total number of triples and nodes and then subtracting a predefined baseline. In this context, the baseline represents the KG before the inclusion of start and stop dates for medication, reflecting the dataset's structure without temporal information. The ?additionalTriples value indicates the number of triples added to the dataset due to the inclusion of this temporal data, offering insight into how the graph has expanded beyond the baseline.

Listing 14- Sparql query for calculating metrics related to the complexity of modelling.

```
SELECT  ?TotalNodes ?Triples  ?additionalTriples ?Redundancy
WHERE {
  {
SELECT (COUNT(*) as ?Triples)
WHERE {
            ?s ?p ?o
    }
}
{
SELECT (COUNT(DISTINCT ?Nodes) AS ?TotalNodes)
WHERE
{
  {
    ?Nodes ?p []
  }
  UNION
  {
    [] ?p ?Nodes
  }
}
    }
    BIND(((?Triples + ?TotalNodes) - (8934 +3916)) AS ?Redundancy)
    BIND((?Triples - 8934) AS ?additionalTriples)
}
```

Table 22 shows the results of evaluating complexity for all the approaches. The table is based on the calculation in Listing 14.

Table 22- The results of complexity per temporal data modelling approaches

Metrics category	Metrics	1. SR	2. RM	3. TG	4. RT	5. SP	6. RS	7.NG

Complexity	Number of nodes	11,636	11,840	19,531	15,137	18,637	10,778	9,345
	Number of triples	43,348	58,853	74,465	101,312	51,195	34,948	28,618
	Additional triple generation	34,414	49,919	65,531	92,378	3,7261	28,643	19,684
	Resource redundancy	42,134	57,843	81,146	103,599	42,982	29,732	25,113

The complexity of modelling in these different approaches is reflected in the number of nodes, triples, additional triple generation, and resource redundancy. Approach TG (3) is the most complex, with the highest number of nodes (19,531) and triples (74,465), indicating a more detailed and possibly granular modelling approach. In contrast, approach NG (7) is the least complex, with the lowest values across all metrics, suggesting a more streamlined model. Approaches SR (1) and RM (2) fall in the middle range, but still involve significant complexity, while RT (4) has the highest overall triple and redundancy generation, likely indicating an extensive use of resource duplication or overlap. Overall, approaches with higher node and triple counts, such as TG and RT, tend to introduce more modelling complexity due to increased resource use and data generation, whereas approaches like NG focus on simplicity with fewer nodes and triples.

To enhance clarity and comprehension of the data, the results have been presented in bar charts, offering a visual representation of the results.

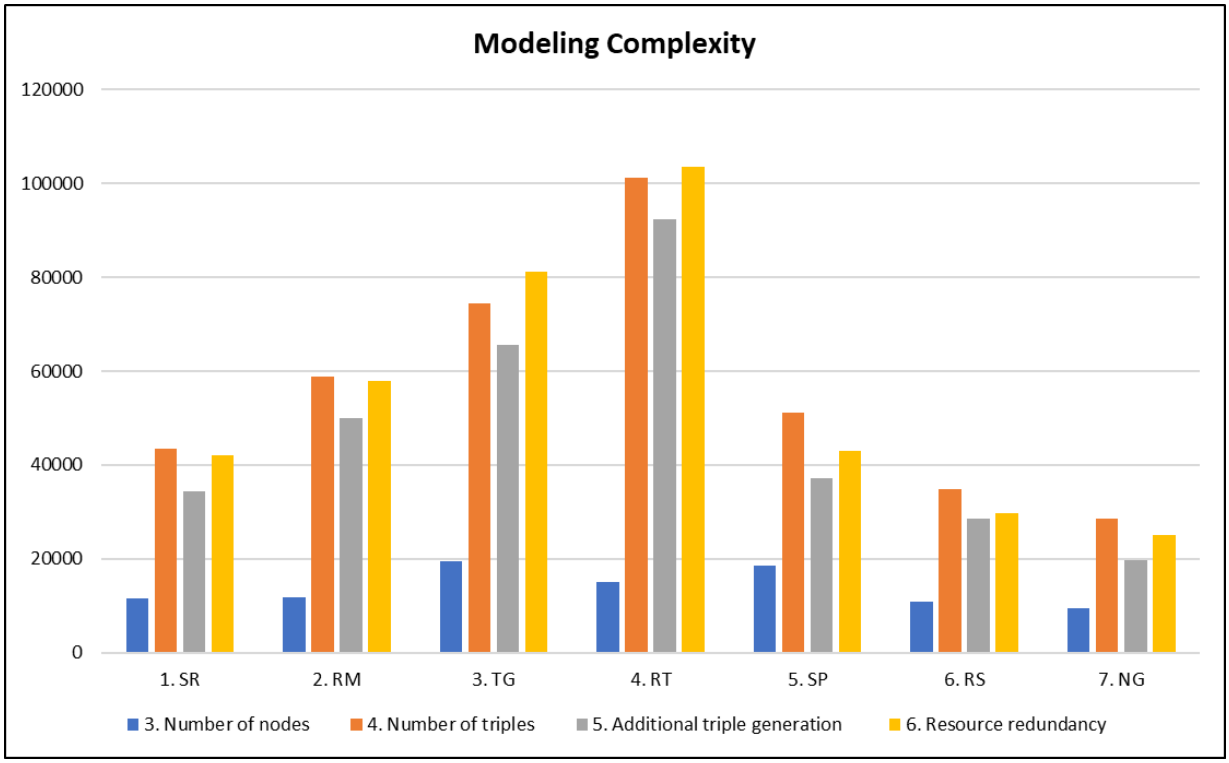


Figure 23- Modelling complexity based on different approaches.

Although the query provides objective metrics such as the number of triples, distinct nodes, and a derived redundancy value, human judgment is essential for interpreting what these numbers mean in the context of knowledge graph complexity. For example:

- A high number of additional triples might indicate richer information or, alternatively, unnecessary verbosity.
- A high redundancy value could reflect useful duplication for context (e.g., inference or provenance) or inefficient modelling practices.
- The total node count may suggest granularity, but understanding whether this improves or hinders usability depends on domain needs and modelling goals.

Ultimately, while the metrics provide a quantitative snapshot, human interpretation is needed to evaluate whether the structure is well-modeled, over-engineered, or lacking in semantic clarity.

6.3.2 Uplifting Metrics Category

Secondly, the uplifting metrics category was considered. This category includes three subcategories: implementation support and standardisation, complexity, and performance.

6.3.2.1 Implementation support and standardisation

Table 23- Implementation support and standardisation-Uplifting category

Metrics category	Metrics	1. SR	2. RM	3. TG	4. RT	5. SP	6. RS	7.NG
Implementation support and standardisation	Uplift engine support	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Uplift adherence to W3C	Yes	Yes	Yes	Yes	Yes	Yes	Yes

6.3.2.2 Complexity

Based on the results of uplifting complexity, TG and RM exhibit the highest uplifting complexity, with 7 and 6 triple maps and 21 term maps each, indicating that these approaches require more extensive mappings to convert data into RDF while preserving temporal aspects. In contrast, RS and NG are less complex, with only 3 triple maps each and fewer term maps (10 and 12, respectively), reflecting a simpler conversion process.

Table 24- Uplifting Complexity results

Metrics category	Metrics	1. SR	2. RM	3. TG	4. RT	5. SP	6. RS	7.NG
Complexity	Number of triple maps	5	6	7	5	4	3	3
	Number of term maps	17	21	21	17	14	10	12

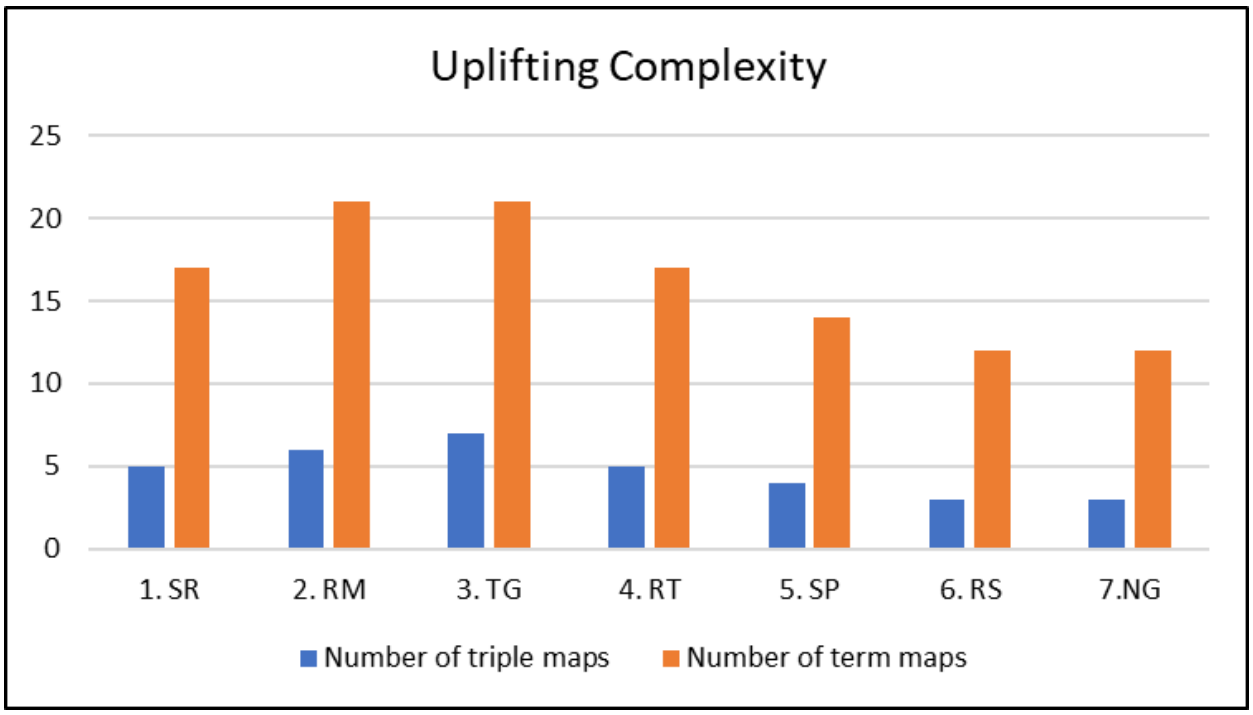


Figure 24- Uplifting complexity per approach

6.3.2.3 Performance

Table 25- RDF data storage space in KB

Metric category	Metric	1. SR	2. RM	3. TG	4. RT	5. SP	6. RS	7.NG
Complexity	RDF data storage space	5,125	6,167	9,471	9,973	3,860	3,542	3,387

The RDF data storage space metric compares the storage requirements of seven different temporal data modelling approaches. Approaches TG (9471) and RT (9973) appear to require the most storage space, suggesting they involve more data and possibly higher complexity. SR (5125) and RM (6167) are in the middle range, while SP (3860), RS (3542), and NG (3387) use less storage space, indicating that these approaches may be more efficient in terms of storage. This metric suggests that TG and RT may involve more data generation, while NG, RS, and SP may be more lightweight in terms of storage needs.

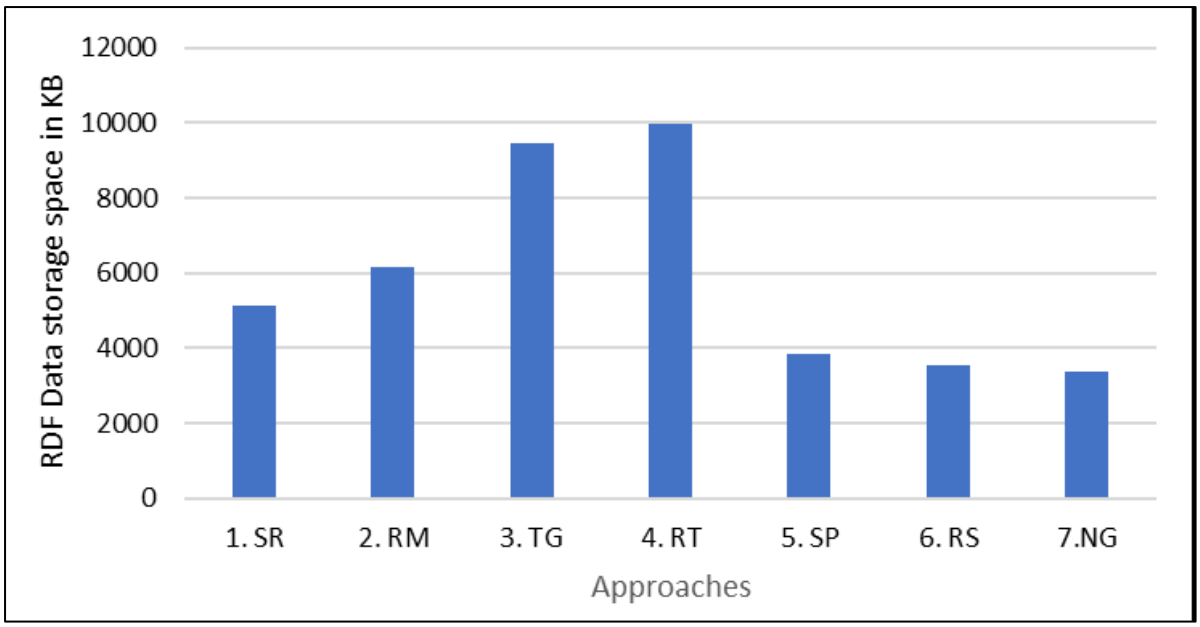


Figure 25- RDF Data storage space in KB per approaches

6.3.3 Querying Metrics Category

Finally, the Querying metrics category was considered. This category includes two subcategories: implementation support and standardisation, and performance.

6.3.3.1 Implementation support and standardisation

In terms of W3C query language support, all the systems fully comply with the standards, as indicated by the "Yes" entries across the board. Similarly, each system provides support for triple store implementation, showing uniformity in their ability to handle RDF triples and structured data storage.

Table 26- Implementation support and standardisation per approaches in Querying metrics category

Metrics category	Metrics	1. SR	2. RM	3. TG	4. RT	5. SP	6. RS	7.NG
Implementation support and standardisation	W3C query language support	Yes	Yes	Yes	Yes	Yes	Yes	Yes
	Triple store support	Yes	Yes	Yes	Yes	Yes	Yes	Yes

6.3.3.2 Performance

In this category, three metrics were assessed includes data load time, query length, and query response time.

Data Load time: The data load time results show that RDF-STAR and Named Graph are the most efficient approaches, with the fastest load times of 1 second, making them

potentially better suited for scenarios requiring quick data initialisation. Standard Reification, RDFT, and Singleton Property fall in the mid-range with 2 seconds, offering moderate performance. On the slower end, RDFMT and Temporal RDF Graph take 3 seconds to load data, which could be a disadvantage in environments that need faster data handling. Overall, RDF-STAR and Named Graph stand out as potentially more efficient for rapid data processing.

Table 27-Data load time per approaches

Metrics category	Metrics	1. SR	2. RM	3. TG	4. RT	5. SP	6. RS	7.NG
Performance	Data load time	2	3	3	2	2	1	1

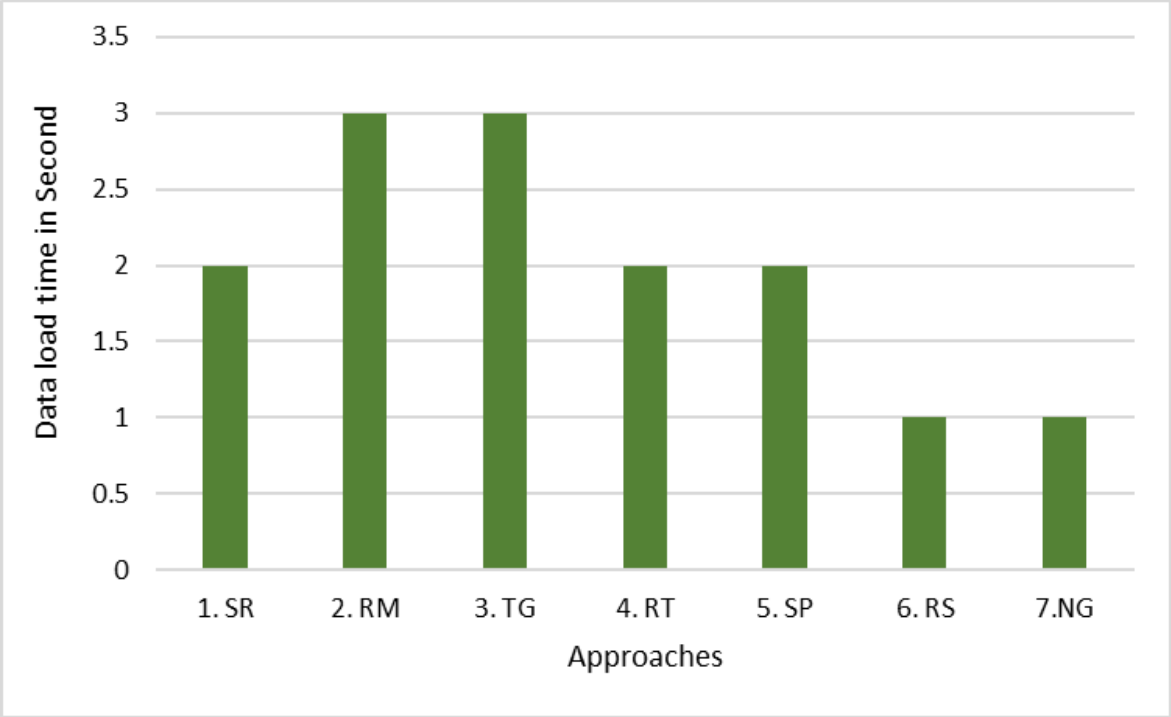


Figure 26- Data load time per approaches.

Query Length: The query length results suggest that Singleton Property and RDF-STAR are potentially more efficient in terms of query size, as they tend to produce shorter queries across most categories. RDFMT and RDFT, on the other hand, generate longer queries, particularly for implicit and ordinal temporal queries, which may indicate higher complexity in query construction. Standard Reification, Temporal RDF Graph, and Named Graph maintain moderate query lengths, with some variation depending on the temporal category. Overall, shorter query lengths may suggest ease of processing and potential efficiency in execution.

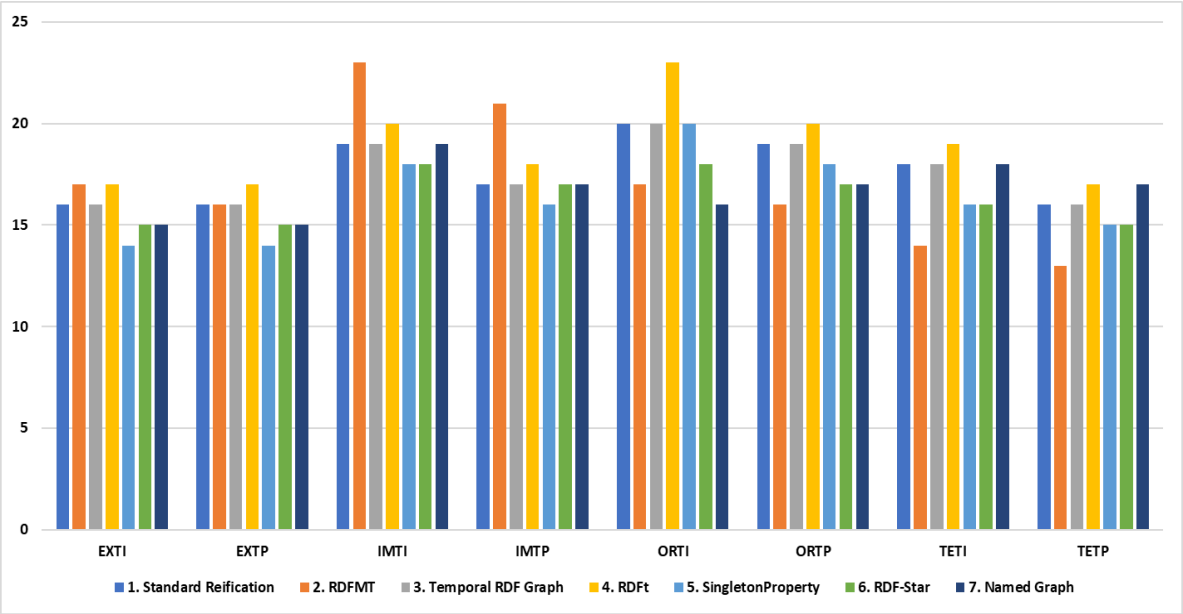


Figure 27- Query length per temporal question categories.

Query response time: Based on the results of query response times for different temporal categories, the performance across most approaches appears similar, particularly for implicit temporal, ordinal temporal, and temporal answer queries. This indicates that the dataset may not be large enough to show major performance differences, as response times in most cases remain consistent, often at 0.1 second. As a result, it is difficult to make definitive decisions on which approach is best solely based on these current numbers, though the data suggests that most methods handle these queries with relative efficiency given the current data volume.

However, certain approaches, such as RDFT, RDF-STAR, and Named Graph, may potentially offer better performance when dealing with larger datasets or more complex queries. Their consistent performance across all temporal categories could make them more scalable in different scenarios. In contrast, Singleton Property shows slower response times, particularly for implicit and ordinal temporal queries, suggesting that it might not be as effective in situations where query speed is crucial. Nonetheless, while the data points to potential differences, more conclusive results would require testing with larger or more complex datasets

Table 28- Query response time per approaches per temporal question categories

APPROACH	EXTI	EXTP	IMTI	IMTP	ORTI	ORTP	TETI	TETP
1. STANDARD REIFICATION	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1
2. RDFMT	0.3	0.1	0.1	0.1	0.1	0.1	0.1	0.1
3. TEMPORAL RDF GRAPH	0.12	0.4	0.1	0.1	0.1	0.1	0.1	0.1
4. RDFT	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
5. SINGLETONPROPERTY	0.6	0.5	2	2	3	3	2	2
6. RDF-STAR	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
7. NAMED GRAPH	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1

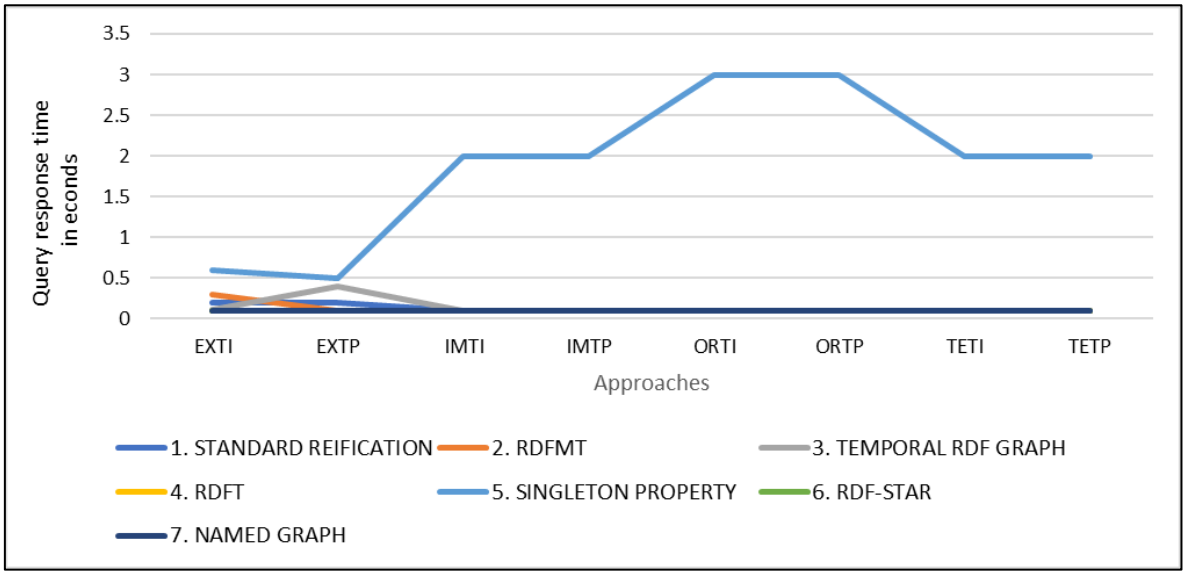


Figure 28- Query response time per approaches per temporal question category

All query response time measurements were conducted on the same system to ensure consistency across all modeling approaches. Units were recorded in milliseconds during data collection and later converted to seconds for reporting purposes. To improve accuracy, each query was executed multiple times (typically 3), and the average response time was calculated. Additionally, all datasets were preloaded into GraphDB prior to testing to eliminate variability caused by data loading or indexing delays. The times presented in the results are the exact values measured during the experiments and were not rounded. The reason many of the time figures appear similar is due to the relatively small dataset size and the consistent performance of the modelling approaches under the controlled testing environment.

6.4 Chapter Summary

In this chapter, a detailed experimental study was conducted to evaluate several temporal data modelling approaches within the context of the FAIRVASC use case, using the TEDME-KG metrics framework. The evaluation spanned multiple phases, including the identification of data modelling approaches, use case definition, knowledge graph creation and deployment, and metrics-based performance assessment. Through this systematic analysis, potential differences were observed in the complexity and standardisation of the approaches. Approaches like TG and RT exhibited relatively higher complexity due to their significant node and triple counts, which could potentially impact scalability and performance, while NG and RS appeared to offer more streamlined models, adhering to W3C standards and demonstrating lower complexity. Despite some advantages in certain areas, none of the methods supported ontology editors, which may limit practical implementation. Query performance varied across approaches, with RDF-STAR and Named Graph emerging as potentially more efficient in terms of data load time and query response times, suggesting they may be better suited for scalable applications. Overall, this chapter addressed Objective 4 of the thesis by evaluating the applicability of the developed metrics framework and exploring the research question: "To what extent can a metrics framework facilitate the evaluation of temporal data modelling approaches in knowledge graphs for Semantic Web researchers?" The study indicated that the metrics framework could be effective in highlighting trade-offs between model complexity, standardisation, and performance, providing useful insights for Semantic Web researchers in selecting suitable temporal modelling approaches.

6.4.1 Comparison with baselines methods

The TEDME-KG framework represents a substantial advancement over existing evaluation approaches for RDF temporal data modelling. Unlike traditional benchmarking tools and isolated modelling techniques, TEDME-KG offers a comprehensive and integrated methodology that evaluates modelling strategies across the entire knowledge graph lifecycle—namely, data modelling, data uplifting, and querying. Existing benchmarks such as RDF Reification Benchmark, StarBench, and WasDiv predominantly focus on system-level performance metrics such as query execution time, storage efficiency, or scalability. While useful for stress-testing RDF systems, these benchmarks do not account for the complexities of representing temporal data or the challenges involved in transforming and

querying it effectively. They provide a narrow view, often overlooking semantic completeness, modelling effort, and user-oriented expressiveness.

Moreover, prior work on RDF temporal modelling approaches—such as the singleton property, named graphs, n-ary relations, and RDF*—often evaluates only syntactic representation or theoretical properties. These approaches typically lack rigorous, quantitative assessment of how well they capture temporal semantics or how usable they are for data engineers. Comparative studies in the literature, such as those by Fu et al., Wang & Tansel, and Hernandez et al., are limited to select performance indicators like the number of triples or response time, and rarely offer a holistic evaluation. Importantly, none of these studies provide a systematic methodology for designing or applying evaluation metrics, nor do they reflect a framework that spans from model design through to query execution.

In contrast, TEDME-KG introduces a robust and structured evaluation process based on a clear hypothesis: that a temporal RDF data modelling approach must be assessed not only by its performance, but also by its modelling expressiveness, transformation complexity, and utility in answering temporally rich queries. It operationalizes this hypothesis by defining a custom set of metrics for each of the three lifecycle stages. For modelling, it evaluates temporal representation capabilities, semantic completeness, and modelling overhead. For data uplift, it introduces metrics like rule complexity, number of mapping rules, and effort required. For querying, TEDME-KG assesses the ability of the model to support different classes of queries, including those requiring temporal reasoning and aggregation.

TEDME-KG further stands out by being the only approach among its peers to offer both a systematic methodology for metric design and a repeatable evaluation process. Its application across multiple modelling approaches—six in total—ensures that the comparison is not only comprehensive but also fair and methodologically grounded. Unlike baseline methods that are often limited to theoretical scenarios or narrowly scoped datasets, TEDME-KG's evaluation is grounded in real-world data and aligned with user expectations, such as clarity, completeness, and usability of temporal knowledge. It enables researchers and practitioners to make informed decisions when selecting or designing temporal RDF models, filling a critical gap in the state of the art.

6.4.2 Experiment Limitations

While the experiments offered valuable insights into the trade-offs among different temporal data modelling approaches, several limitations should be acknowledged. The dataset used was relatively small, potentially masking performance differences—particularly in query response time—among the approaches. Moreover, the lack of ontology editor support across all methods presents a practical barrier to real-world adoption, regardless of theoretical strengths in standardisation or complexity. Although RDF-STAR and Named Graph consistently showed lower complexity and better performance, further validation with larger, more diverse datasets is necessary to confirm their scalability. Additionally, while higher redundancy and triple counts in approaches like TG and RT may reflect richer semantics, they may also signal inefficiency or over-engineering, depending on specific use cases. These findings highlight the importance of aligning model selection not just with technical metrics but also with implementation context and intended data use. The study ultimately demonstrates the usefulness of the TEDME-KG framework for revealing strengths and trade-offs, yet it also suggests a need for complementary qualitative assessments to guide informed modelling decisions.

Chapter 7 Conclusion

This chapter summarises the key findings and achievements of the research. It includes a summary of how the four main research objectives were met, highlighting the development of the TEDME-KG metrics framework, the creation of a quality assessment framework for healthcare data sources, and the systematic review of temporal data modelling approaches. Additionally, it discusses the impact and practical applications of these contributions, particularly in healthcare settings. The chapter also outlines potential future research directions and concludes with final remarks on the theoretical and practical advancements made through this research.

7.1 Achievement of Research Objectives

This thesis set out to achieve four key research objectives to address the research question:

"To what extent can a metrics framework facilitate the evaluation of temporal data modelling approaches in knowledge graphs for Semantic Web researchers?"

These objectives were designed to answer the research question thereby contributing to the field of temporal data modelling evaluation in KGs. Below is discussion of how each objective was met:

Achievement of Research Objective 1 (RO1)

RO1: Identify the problem context and use case motivation by:

- Development of a framework to assess the quality of data sources in healthcare settings and investigate the importance of time and temporal data as a criterion for the quality of data and data sources.
- Exploring the FAIRVASC use case and identifying the associated challenges.

RO1 was achieved by developing a framework to assess the quality of data sources in healthcare settings, investigating the importance of time and temporal data as a criterion for the quality of data and data sources and exploring approaches for modelling temporal data in healthcare.

The development of a quality assessment framework was initiated with an extensive literature review, aimed at gathering existing frameworks and criteria for assessing the quality of data sources in healthcare. This review involved comprehensive searches through academic databases, journals, and grey literature, ensuring a thorough understanding of current practices and standards. The insights gained from this review were instrumental in creating a novel framework tailored specifically for healthcare settings. To ensure the framework's practical relevance and robustness, it was refined through expert advisory group meetings. This iterative process ensured that the framework was both theoretically sound and practically applicable, addressing real-world needs and challenges. Time was identified as a criterion important for the quality of data and data sources in healthcare.

Finally, an exploration of the FAIRVASC use case, along with the challenges associated with temporal data modelling, was conducted through participation in several meetings with the FAIRVASC research group.

By addressing the dual focus on quality assessment and temporal data modelling, the research ensured that both the foundational and applied aspects of the problem context were comprehensively understood, leading to solutions that are both theoretically sound and practically applicable.

Achievement of Research Objective 2 (RO2)

- RO2: Identify approaches for modelling temporal data in KGs and healthcare and investigate methods for evaluating data modelling approaches in KGs.

RO2 was achieved by conducting state-of-the-art reviews to identify approaches for modelling temporal data in KGs and methods for evaluating data modelling approaches in KGs. This objective was achieved through a systematic and detailed review process.

The identification of temporal data modelling approaches began with an exhaustive systematic review of the literature. This review focused on various approaches employed in modelling temporal data in KGs, categorising them based on their semantics, time

properties, scope of application and use case. The comprehensive analysis provided a clear and detailed understanding of the current state of the art in temporal data modelling in KGs.

The exploration of temporal data modelling approaches in healthcare was conducted using a literature review, which provided an understanding of existing approaches and applications in modelling temporal data in healthcare. This review investigated the approaches and applications of temporal data modelling in healthcare

In addition, another state of art review was conducted to identify existing methods for evaluating data modelling approaches in KGs.

These reviews offered an understanding of both the existing temporal data modelling approaches and their evaluation methods in KGs.

Achievement of Research Objective 3 (RO3)

- RO3: Design a metrics framework for evaluating temporal data modelling approaches in KGs.

RO3 was achieved by the development of the TEDME-KG metrics framework through a rigorous and iterative process involving several key steps to ensure its robustness and applicability.

The framework development process began with the adoption of the Goal Question Metric (GQM) approach, which provided a structured methodology for defining high-level goals, formulating specific questions to assess these goals, and identifying relevant metrics to answer the questions. This structured approach ensured that the framework was comprehensive and systematically organised, covering all essential aspects of temporal data modelling. The GQM approach was particularly effective in aligning the framework with both theoretical considerations and practical requirements, ensuring that each metric was relevant and actionable.

A detailed literature review was conducted to identify existing metrics and methodologies related to temporal data modelling in knowledge graphs. This review was complemented by consultations with domain experts to gather insights and validate the relevance of the identified metrics. These experts provided invaluable feedback on the practical aspects of the metrics, ensuring that the framework was not only theoretically sound but also practically

applicable. This iterative process of literature review and expert consultation allowed for the continuous refinement of the framework, incorporating the latest advancements and best practices in the field.

To support the practical application of the metrics framework, comprehensive documentation was developed. This documentation includes detailed definitions and examples, making the framework accessible and usable by researchers and practitioners. The documentation ensures that users can easily understand and apply the metrics framework to various temporal data modelling approaches. It provides clear instructions on how to evaluate different aspects of temporal data modelling, ensuring consistency and reliability in the evaluation process.

The combination of a structured development approach, thorough literature review, expert input, and comprehensive documentation ensured that the TEDME-KG metrics framework is both robust and user-friendly. The framework is capable of providing meaningful evaluations of temporal data modelling techniques, addressing key aspects such as accuracy, scalability, and usability. By offering a systematic and comprehensive evaluation tool, the TEDME-KG metrics framework significantly advances the field of temporal data modelling in knowledge graphs, providing researchers and practitioners with the necessary tools to assess and improve their methodologies effectively.

Achievement of Research Objective 4 (RO4)

- RO4: Evaluate the developed metrics framework in terms of its appropriateness, applicability, and general usability as perceived by Semantic Web researchers.

RO4 was achieved through a comprehensive evaluation strategy:

The evaluation began with conducting a survey, where the TEDME-KG metrics framework was assessed by Semantic Web researchers. The evaluation focused on assessing the framework's appropriateness, design, and general usability. The experts' feedback was crucial in identifying strengths and areas for improvement, ensuring that the framework met the needs of its users. The feedback from expert evaluations was used to iteratively refine the framework.

In addition to expert evaluations, the TEDME-KG metrics framework was applied to a real-world use case in healthcare, to evaluate its practical applicability. These evaluations ensured that the TEDME-KG metrics framework was both theoretically sound and practically applicable, capable of providing valuable evaluations of temporal data modelling approaches.

7.2 Contributions

The contributions of this thesis include theoretical advancements and practical applications. The major contribution is the development of the TEDME-KG metrics framework, designed to evaluate temporal data modelling approaches in RDF-based knowledge graphs. This framework represents a structured and systematic approach to assess various aspects of temporal data modelling including modelling, uplifting, and querying. It addresses a critical gap in the current literature by providing a comprehensive tool that facilitates better evaluation of temporal data modelling approaches in KGs.

The TEDME-KG metrics framework was developed through a rigorous and iterative process, combining a structured methodology with expert input and thorough literature review. The framework adopts the Goal Question Metric (GQM) approach, which ensures a clear alignment of high-level goals, specific questions, and relevant metrics. This structured approach allows for a detailed and organised assessment of temporal data modelling techniques, making the framework robust and comprehensive. By grounding the framework in both theoretical considerations and practical requirements, the GQM approach ensures that each metric is relevant and actionable, addressing key aspects of temporal data modelling effectively.

A literature review was conducted to identify existing metrics and methodologies related to temporal data modelling in knowledge graphs. This review was complemented by consultations with domain experts to gather insights and validate the relevance of the identified metrics. These experts provided invaluable feedback on the practical aspects of the metrics, ensuring that the framework was not only theoretically sound but also practically applicable. This iterative process of literature review and expert consultation allowed for the continuous refinement of the framework, incorporating the latest advancements and best practices in the field. By integrating expert insights, the framework became more robust and

user-friendly, capable of providing meaningful evaluations of temporal data modelling techniques.

The practical applicability of the TEDME-KG metrics framework is further enhanced by comprehensive documentation, which includes detailed definitions and examples. This documentation ensures that users can easily understand and apply the metrics framework to various temporal data modelling approaches. Clear instructions on how to evaluate different aspects of temporal data modelling ensure consistency and reliability in the evaluation process. The emphasis on critical aspects such as accuracy, scalability, and usability ensures that the framework provides meaningful evaluations that can guide the improvement of temporal data modelling techniques. This, in turn, enhances the overall quality and performance of knowledge graphs, which are essential tools in various domains, including healthcare, finance, and information technology.

In addition to the TEDME-KG metrics framework, this thesis made several other significant contributions. One notable contribution is the development of a comprehensive framework to assess the quality of data sources in healthcare settings. This framework includes eight parent themes and 22 sub-themes, offering a detailed and nuanced evaluation tool. It assists healthcare data users in identifying and evaluating multiple quality aspects, ensuring data reliability and usability. The framework also serves as a valuable resource for policymakers, health information standard organisations, and data development teams, guiding the establishment of standards, guidelines, and legislation regarding data sources in healthcare settings. By addressing the quality of data sources, this framework helps to ensure that healthcare data is reliable and useful, ultimately improving patient care and healthcare outcomes.

Another contribution is the creation of a categorised list of approaches and applications for temporal data modelling in knowledge graphs. This list, developed through a systematic review process, provides a clear and organised reference for researchers and practitioners. By highlighting trends and gaps in current research, this list guides future investigations in temporal data modelling within knowledge graphs, helping to advance the field and promote the development of innovative solutions.

Empirical evaluation results for selected temporal data modelling approaches were also presented in this thesis, offering valuable insights into their performance and suitability in

different contexts. These results provide a foundation for further research and development, contributing to the ongoing advancement of temporal data modelling techniques. By systematically evaluating and comparing different approaches, the research identified strengths and weaknesses, offering practical recommendations for improving temporal data modelling in knowledge graphs.

In summary, the TEDME-KG metrics framework stands out as a major contribution of this thesis, providing a robust, systematic, and practical tool for evaluating temporal data modelling approaches in RDF-based knowledge graphs. Alongside this, the additional frameworks and categorised lists significantly enrich the field, providing essential tools and insights that enhance data quality assessment and modelling methodologies in healthcare and other domains. These contributions collectively address critical gaps in the literature, offering innovative solutions and practical tools that advance the state of the art in temporal data modelling and data quality assessment. By facilitating better decision-making and implementation of temporal data models, this thesis contributes to the improvement of knowledge graphs and their applications across various fields.

7.3 Impact and Uptake

The contributions of this thesis have implications for both theoretical advancements and practical applications in the fields of temporal data modelling and knowledge graphs.

One of the most impactful aspects of this thesis is its application in real-world use case, particularly in the healthcare domain. The TEDME-KG metrics framework has been instrumental in evaluating different temporal data modelling approaches, which in turn, has guided the extension and improvement of the FAIRVASC ontology. This ontology benefited from a systematic evaluation of different temporal data modelling approaches, in modelling temporal relationships, such as medication start and stop. The TEDME-KG metrics framework evaluates temporal data modelling approaches across three categories: modelling, uplifting, and querying. Additionally, during the design of different KGs based on different temporal data modelling approaches, new solutions were proposed to the FAIRVASC team for modelling medication start date and stop date.

In healthcare settings, the quality assessment framework developed in this thesis provides a tool for evaluating data sources. With its eight parent themes and 22 sub-themes, the

framework helps healthcare data users identify and assess various quality aspects, ensuring data reliability and usability. This is particularly crucial in healthcare, where data quality directly impacts patient care and outcomes. Policymakers and health information standard organisations can utilise this framework to establish standards, guidelines, and legislation, thereby improving the quality and consistency of healthcare data sources. Additionally, healthcare data development teams can use the framework to consider different aspects of a data source prior to its design and implementation, ensuring that new data sources meet high-quality standards from the outset. The data source quality assessment framework was published in a prestigious Journal of the American Medical Informatics Association (JAMIA). It was also presented at the annual Health Information and Quality Authority (HIQA) PhD showcase event for all the HIQA staff. Additionally, Science Foundation Ireland (SFI) featured a post on their website and conducted an interview with the researchers involved in this study.

In summary, the work presented in this thesis has not only advanced theoretical understanding in the field of temporal data modelling but also provided practical tools that have been successfully applied in real-world scenarios. The frameworks and methodologies developed offer a strong foundation for future research and development, ensuring that temporal data models continue to evolve and meet the complex needs of various applications. This research thus represents a significant step forward in the quest for more effective and reliable data modelling techniques, with broad implications for improving data quality and utility in numerous fields.

7.4 Future Work

While this thesis has made significant contributions, there are several areas where future research can further expand and enhance the findings:

Further Validation and Refinement of TEDME-KG Metrics Framework: Additional case studies and applications in diverse domains are essential to further validate and refine the TEDME-KG metrics framework. These studies can provide more comprehensive insights into the framework's effectiveness and identify any remaining gaps or areas for improvement. By applying the metrics framework to a wider variety of temporal data modelling scenarios, researchers can gather a more extensive range of feedback. This

iterative process of validation and refinement will help ensure that the framework remains robust, and applicable across different contexts.

Moreover, exploring the integration of the TEDME-KG metrics framework with other data modelling and evaluation tools could enhance its usability and effectiveness. By creating synergies with existing tools, the framework can become more accessible and practical for researchers and practitioners. This integration might involve developing plugins or modules that allow seamless interaction with popular data modelling software, facilitating easier adoption and broader usage of the framework. Such enhancements could make the framework a standard tool in the evaluation of temporal data modelling approaches, thereby increasing its impact and utility.

Advancing Temporal Data Modelling Techniques: Future research could also focus on advancing temporal data modelling techniques themselves. While the TEDME-KG metrics framework provides a robust tool for evaluating existing approaches, there is always room for innovation and improvement in the methodologies used to model temporal data. Researchers could explore new approaches to enhance the accuracy, efficiency, and effectiveness of temporal data models.

Interdisciplinary Applications and Collaborations: Interdisciplinary applications and collaborations could also play a crucial role in future research. Temporal data modelling is relevant to numerous fields, including environmental science, finance and social sciences. Collaborating with experts in these areas could provide new perspectives and drive the development of tailored solutions for specific applications. Such collaborations could also facilitate the transfer of knowledge and best practices across disciplines, enriching the research community and fostering innovation.

Policy and Ethical Considerations: Finally, future research should also consider the policy and ethical implications of temporal data modelling. As data-driven decision-making becomes increasingly prevalent, it is essential to ensure that data models are used responsibly and ethically. Researchers could investigate the ethical challenges associated with temporal data modelling, such as privacy concerns, data biases, and the potential for misuse. Developing guidelines and best practices for the ethical use of temporal data models could help address these challenges and promote responsible data science.

In summary, while this thesis has made substantial contributions to the field of temporal data modelling, there are numerous opportunities for future research to expand and enhance these findings. By validating and refining the TEDME-KG metrics framework, advancing data modelling approaches, fostering interdisciplinary collaborations, and addressing policy and ethical considerations, researchers can continue to push the frontiers of knowledge and practice in this area.

7.5 Final Remarks

This thesis has made contributions to the field of temporal data modelling in KGs, addressing challenges and advancing the state of the art through the development of the TEDME-KG metrics framework. The comprehensive approach taken in this research, which included systematic literature reviews, expert consultations, and rigorous evaluations, has resulted in a robust framework for evaluating temporal data modelling approaches. The TEDME-KG metrics framework provides a structured way for evaluating key aspects such as modelling, uplifting, and querying, ensuring that temporal data modelling approaches are comparable across different categories.

The practical applicability of the framework was demonstrated through its use in the FAIRVASC case study, where it facilitated the extension and improvement of the FAIRVASC ontology. This real-world application highlighted the framework's capacity to drive enhancements in temporal data management. The success of the TEDME-KG metrics framework in this context underscores its potential to serve as a valuable tool for researchers and practitioners working with temporal data modelling approaches.

Beyond the framework itself, this thesis has also provided valuable resources in the form of a categorised list of temporal data modelling approaches. These contributions offer a foundation for ongoing research and practice, guiding the development and evaluation of temporal data models across various domains. The systematic organisation of temporal data modelling approaches facilitates the identification of suitable methods for specific applications, ensuring that temporal data models are effective and well-suited to their respective contexts. Moreover, the development of data source quality assessment framework in healthcare settings, provides

In conclusion, this thesis represents the TEDME-KG metrics framework and associated contributions which have the potential to drive innovation and improvement in temporal data management, offering substantial benefits to both academia and industry. By addressing critical gaps in the literature and providing practical solutions, this research has laid a strong foundation for future advancements in the field. The work presented here not only advances theoretical understanding but also provides practical tools that can be successfully applied in real-world scenarios. This research thus marks a significant milestone in the ongoing effort to enhance the area of temporal data modelling in KGs.

References

1. Zou, X., *A Survey on Application of Knowledge Graph.* Journal of Physics: Conference Series, 2020. **1487**: p. 012016.
2. Yan, J., et al., *A retrospective of knowledge graphs.* Frontiers of Computer Science, 2018. **12**(1): p. 55-74.
3. Wang, J., et al., *3DRTE: 3D Rotation Embedding in Temporal Knowledge Graph.* IEEE Access, 2020. **8**: p. 207515-207523.
4. Chen, Z., et al., *Knowledge Graph Completion: A Review.* IEEE Access, 2020. **8**: p. 192435-192456.
5. Fensel, D., et al., *Introduction: What Is a Knowledge Graph?*, in *Knowledge Graphs: Methodology, Tools and Selected Use Cases*, D. Fensel, et al., Editors. 2020, Springer International Publishing: Cham. p. 1-10.
6. Hogan, A., et al., *Knowledge graphs.* Synthesis Lectures on Data, Semantics, and Knowledge, 2021. **12**(2): p. 1-257.
7. Daniele, D.A., et al., *RSP-QL Semantics: A Unifying Query Model to Explain Heterogeneity of RDF Stream Processing Systems.* International Journal on Semantic Web and Information Systems (IJSWIS), 2014. **10**(4): p. 17-44.
8. Zou, L. and M.T. Özsu, *Graph-Based RDF Data Management.* Data Science and Engineering, 2017. **2**(1): p. 56-70.
9. YounHee, K., K. ByungGon, and L. HaeChull. *The index organizations for RDF and RDF schema*. in *2006 8th International Conference Advanced Communication Technology*. 2006.
10. Wang, H.-T. and A. Tansel, *Temporal Extensions to RDF.* Journal of Web Engineering, 2019. **18**: p. 125-168.
11. Ni, R., et al. *Specific Time Embedding for Temporal Knowledge Graph Completion*. 2020.
12. Zhang, X., J. Qian, and Q. Wu. *Research on the solution to redundancy of temporal data*. in *2010 International Conference on Computer Application and System Modeling (ICCASM 2010)*. 2010.
13. Xu, Y., et al., *Temporal Knowledge Graph Reasoning with Historical Contrastive Learning.* Proceedings of the AAAI Conference on Artificial Intelligence, 2023. **37**(4): p. 4765-4773.
14. Gutierrez, C., C.A. Hurtado, and A. Vaisman, *Introducing time into RDF.* IEEE Transactions on Knowledge and Data Engineering, 2007. **19**(2): p. 207-218.
15. Orlandi, F., D. Graux, and D. O'Sullivan. *Benchmarking RDF Metadata Representations: Reification, Singleton Property and RDF∗*. in *Proceedings - 2021 IEEE 15th International Conference on Semantic Computing, ICSC 2021*. 2021.
16. Zhang, F., et al., *RDF for temporal data management – a survey.* Earth Science Informatics, 2021. **14**(2): p. 563-599.
17. Sen, S., et al., *State-of-the-Art Approaches for Meta-Knowledge Assertion in the Web of Data.* IETE Technical Review, 2021. **38**(6): p. 672-709.
18. Ribeiro, A., A. Silva, and A.R.d. Silva, *Data Modeling and Data Analytics: A Survey from a Big Data Perspective.* Journal of Software Engineering and Applications, 2015. **Vol.08No.12**: p. 18.
19. Poh, N., S. Tirunagari, and D. Windridge. *Challenges in designing an online healthcare platform for personalised patient analytics*. in *2014 IEEE Symposium on Computational Intelligence in Big Data (CIBD)*. 2014.
20. Iroju, O., et al., *Interoperability in healthcare: benefits, challenges and resolutions.* International Journal of Innovation and Applied Studies, 2013. **3**(1): p. 262-270.

21. Jansen-van der Weide, M.C., et al., *Rare disease registries: potential applications towards impact on development of new drug treatments.* Orphanet Journal of Rare Diseases, 2018. **13**(1): p. 154.
22. Boulanger, V., et al., *Establishing Patient Registries for Rare Diseases: Rationale and Challenges.* Pharmaceutical Medicine, 2020. **34**(3): p. 185-190.
23. Olaronke, I., et al., *Interoperability in Healthcare: Benefits, Challenges and Resolutions.* International Journal of Innovation and Applied Studies, 2013. **3**: p. 2028-9324.
24. Nony, P., et al., *A methodological framework for drug development in rare diseases.* Orphanet Journal of Rare Diseases, 2014. **9**(1): p. 164.
25. Yaman, B., et al., *Towards A Rare Disease Registry Standard: Semantic Mapping of Common Data Elements Between FAIRVASC and the European Joint Programme for Rare Disease.*
26. McGlinn, K., et al., *FAIRVASC: A semantic web approach to rare disease registry integration.* Computers in Biology and Medicine, 2022. **145**.
27. Keeney, R.L. and R.S. Gregory, *Selecting attributes to measure the achievement of objectives.* Operations Research, 2005. **53**(1): p. 1-11.
28. Booth, A., *What research studies do practitioners actually find useful?* Health Information & Libraries Journal, 2004. **21**(3): p. 197-200.
29. Wu, D., H.-T. Wang, and A.U. Tansel, *A survey for managing temporal data in RDF.* Information Systems, 2024. **122**: p. 102368.
30. Berners-Lee, T., J. Hendler, and O. Lassila, *The Semantic Web: A New Form of Web Content That is Meaningful to Computers Will Unleash a Revolution of New Possibilities.* ScientificAmerican.com, 2001.
31. De Giacomo, G., et al., *Using Ontologies for Semantic Data Integration*, in *A Comprehensive Guide Through the Italian Database Research Over the Last 25 Years*, S. Flesca, et al., Editors. 2018, Springer International Publishing: Cham. p. 187-202.
32. McKay, S., et al., *Metric Development for Environmental Benefits Analysis.* 2010.
33. Caldiera, V.R.B.G. and H.D. Rombach, *The goal question metric approach.* Encyclopedia of software engineering, 1994: p. 528-532.
34. Benchimol, E.I., et al., *The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement.* PLoS Med, 2015. **12**(10): p. e1001885.
35. Cole, A.P., D.F. Friedlander, and Q.-D. Trinh, *Secondary data sources for health services research in urologic oncology.* Urologic Oncology: Seminars and Original Investigations, 2018. **36**(4): p. 165-173.
36. Gliklich, R.E., M.B. Leavy, and N.A. Dreyer, *Tools and Technologies for Registry Interoperability, Registries for Evaluating Patient Outcomes: A User's Guide, Addendum 2 [Internet].* 2019.
37. HIQA. *What you should know about Data Quality- A guide for health and social care staff*. [cited 2021 Last Access 31/08/2021]; Available from: https://www.hiqa.ie/sites/default/files/2017-01/Data-Quality-Guide.pdf.
38. HIQA. *Guidance on a data quality framework for health and social care*. 2018 Last Access 31/08/2021]; Available from: https://www.hiqa.ie/sites/default/files/2018-10/Guidance-for-a-data-quality-framework.pdf.
39. Kapsner, L.A., et al., *Moving Towards an EHR Data Quality Framework: The MIRACUM Approach.* Stud Health Technol Inform, 2019. **267**: p. 247-253.
40. Almutiry, O., et al. *Toward a framework for data quality in cloud-based health information system*. 2013.
41. Togneri, R., et al. *Foundations of Data Quality Assurance for IoT-based Smart Applications*. in *2019 IEEE Latin-American Conference on Communications (LATINCOM)*. 2019.

42. Information, C.I.f.H. *The CIHI data quality framework*. 2009 Last Access 08/01/2022]; Available from: https://secure.cihi.ca/free_products/dq-data_quality_framework_2009_en.pdf.
43. Health, A. *Data quality framework*. Last Access 31/08/2021]; Available from: https://docplayer.net/6586457-Dgd14-006-act-health-data-quality-framework.html.
44. *Australian Bureau of Statistics, ABS data quality framework*. 2009 Last Access 08/01/2022]; Available from: https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Quality:+The+ABS+data+quality+framework.
45. Mavrogiorgou, A., et al., *Analyzing data and data sources towards a unified approach for ensuring end-to-end data and data sources quality in healthcare 4.0.* Computer Methods and Programs in Biomedicine, 2019. **181**: p. 104967.
46. *Canadian Institute for Health Information*

-*CIHI's Data Source Assessment Tool*. 2017 Last Access 07/02/2023; Available from: https://secure.cihi.ca/estore/productFamily.htm?pf=PFC3554&lang=en&media=0.
47. *Australian Institute of Health and Welfare -An AIHW framework for assessing data sources for population health monitoring*. 2014 Last Access 08/01/2022]; Available from: https://www.aihw.gov.au/reports/health-care-quality-performance/an-aihw-framework-for-assessing-data-sources-for-p/contents/table-of-contents.
48. Denyer, D., D. Tranfield, and J. van Aken, *Developing Design Propositions Through Research Synthesis.* Organization Studies - ORGAN STUD, 2008. **29**: p. 393-413.
49. *Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia*. [cited 2021 31 July 2021]; Available from: https://www.covidence.org/.
50. Microsoft, C. *Microsoft Excel*. 2013 08/01/2022]; 2013 (15.0):[Available from: https://office.microsoft.com/excel.
51. HIQA. *Information management standards for national health and social care data collection*. 2017 Last Access 07/09/2021]; Available from: https://www.hiqa.ie/sites/default/files/2017-02/Information-management-standards-for-national-health-and-social-care-data-collections.pdf.
52. Comission, E. *Ethics guidelines for trustworthy AI*. 2019 Last Access 08/09/2024; Available from: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.
53. Böhlen, M.H., et al., *Temporal data management–an overview.* European Business Intelligence and Big Data Summer School, 2017: p. 51-83.
54. Gong, C., et al., *Causal discovery from temporal data: An overview and new perspectives.* ACM Computing Surveys, 2024. **57**(4): p. 1-38.
55. Kahn, M.G., et al., *A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data.* EGEMS (Wash DC), 2016. **4**(1): p. 1244.
56. Salama, A.D., *Relapse in Anti-Neutrophil Cytoplasm Antibody (ANCA)–Associated Vasculitis.* Kidney International Reports, 2020. **5**(1): p. 7-12.
57. Angles, R. and C. Gutierrez, *Survey of graph database models.* ACM Computing Surveys (CSUR), 2008. **40**(1): p. 1-39.
58. Silberschatz, A., H.F. Korth, and S. Sudarshan, *Data models.* ACM Computing Surveys (CSUR), 1996. **28**(1): p. 105-108.
59. McGee, W.C., *On user criteria for data model evaluation.* ACM Transactions on Database Systems (TODS), 1976. **1**(4): p. 370-387.
60. Coronel, C. and S. Morris, *Database systems: design, implementation, & management*. 2016: Cengage Learning.
61. Madkour, M., D. Benhaddou, and C. Tao, *Temporal data representation, normalization, extraction, and reasoning: A review from clinical domain.* Comput Methods Programs Biomed, 2016. **128**: p. 52-68.

62. Kumar, S. and R. Rishi. *A relative analysis of modern temporal data models*. in *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. 2016.
63. Combi, C., G. Cucchi, and F. Pinciroli, *Applying object-oriented technologies in modeling and querying temporally oriented clinical databases dealing with temporal granularity and indeterminacy.* IEEE Trans Inf Technol Biomed, 1997. **1**(2): p. 100-27.
64. Combi, C., F. Pinciroli, and G. Pozzi. *Temporal clinical data modeling and implementation for PTCA patients in an OODBMS environment*. in *Computers in Cardiology 1994*. 1994.
65. Kahn, M.G. and K.A. Marrs, *Creating temporal abstractions in three clinical information systems.* Proc Annu Symp Comput Appl Med Care, 1995: p. 392-6.
66. Terenziani, P., et al., *Extending temporal databases to deal with telic/atelic medical data.* Artif Intell Med, 2007. **39**(2): p. 113-26.
67. Adlassnig, K.P., et al., *Temporal representation and reasoning in medicine: Research directions and challenges.* p. 101-13.
68. *IEEE Xplore digital library*. Last Access 31 July 2021]; Available from: https://ieeexplore.ieee.org/Xplore/home.jsp.
69. Guarino, N., D. Oberle, and S. Staab, *What Is an Ontology?* 2009. p. 1-17.
70. Telnarova, Z. *Relational database as a source of ontology creation*. in *Proceedings of the International Multiconference on Computer Science and Information Technology*. 2010.
71. Man, L., D. Xiao-Yong, and W. Shan. *Learning ontology from relational database*. in *2005 International Conference on Machine Learning and Cybernetics*. 2005.
72. Chen, P., *Entity-Relationship Modeling: Historical Events, Future Trends, and Lessons Learned*, in *Software Pioneers: Contributions to Software Engineering*, M. Broy and E. Denert, Editors. 2002, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 296-310.
73. Yan, J., et al., *A retrospective of knowledge graphs.* Frontiers of Computer Science, 2016.
74. Stear, B.J., et al., *Petagraph: A large-scale unifying knowledge graph framework for integrating biomolecular and biomedical data.* Scientific Data, 2024. **11**(1): p. 1338.
75. Kitchenham, B. *Procedures for Performing Systematic Reviews*. 2004.
76. *ACM Digital Library*. Last Access 31 July 2021]; Available from: https://dl.acm.org/.
77. *ScienceDirect*. Last Access 31 July 2021]; Available from: https://www.sciencedirect.com/.
78. Zhu, L., et al., *StRDFS: Spatiotemporal Knowledge Graph Modeling.* IEEE Access, 2020. **8**: p. 129043-129057.
79. Haixia, L., *A New query method for the temporal RDF Model RDFMT Based on SPARQL*. 2021, Association for Computing Machinery: Chongqing, China. p. Article 24.
80. Sen, S., et al., *RDFM: An alternative approach for representing, storing, and maintaining meta-knowledge in web of data.* Expert Systems with Applications, 2021. **179**.
81. Chekol, M.W., et al. *Marrying uncertainty and time in knowledge graphs*. 2017.
82. Ramos, M.R., *Biography Light Ontology: An open vocabulary for encoding biographic texts.* Technical Report, 2009.
83. Dong, H., S. Yu, and Y. Jiang. *A Protégé Plug-In for Modeling Dimensional Ontology*. 2009.
84. Motik, B., *Representing and querying validity time in RDF and OWL: A logic-based approach.* Journal of Web Semantics, 2012. **12-13**: p. 3-21.
85. Tao, C., H.R. Solbrig, and C.G. Chute, *CNTRO 2.0: A Harmonized Semantic Web Ontology for Temporal Relation Inferencing in Clinical Narratives.* AMIA Jt Summits Transl Sci Proc, 2011. **2011**: p. 64-8.
86. Tappolet, J. and A. Bernstein, *Applied temporal RDF: Efficient temporal querying of rdf data with SPARQL*. 2009. p. 308-322.
87. Grandi, F. *Multi-temporal RDF ontology versioning*. 2009.
88. Mazzieri, M. and A.F. Dragoni. *Labeled RDF*. 2010.

89. Krieger, H.U. *A temporal extension of the Hayes and ter Horst entailment rules for RDFS and OWL*. 2011.
90. Dylla, M., M. Sozio, and M. Theobald. *Resolving temporal conflicts in inconsistent RDF knowledge bases*. 2011.
91. Kalayci, E.G., et al., *Ontology–based access to temporal data with Ontop: A framework proposal.* Int. J. Appl. Math. Comput. Sci., 2019. **29**(1): p. 17–30.
92. Omelayenko, B. *RDFT: A mapping meta-ontology for business integration*. in *Proc. of the Workshop on Knowledge Transformation for the Semantic Web at the 15th European Conference on Artificial Intelligence (KTSW2002)*. 2002. Citeseer.
93. Nguyen, V., O. Bodenreider, and A. Sheth, *Don't like RDF reification? making statements about statements using singleton property*, in *Proceedings of the 23rd international conference on World wide web*. 2014, Association for Computing Machinery: Seoul, Korea. p. 759–770.
94. Pugliese, A., O. Udrea, and V.S. Subrahmanian. *Scaling RDF with time*. 2008.
95. Perry, M., F. Hakimpour, and A. Sheth, *Analyzing theme, space, and time: an ontology-based approach*, in *Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems*. 2006, Association for Computing Machinery: Arlington, Virginia, USA. p. 147–154.
96. Udrea, O., D.R. Recupero, and V.S. Subrahmanian, *Annotated RDF.* ACM Transactions on Computational Logic, 2010. **11**(2).
97. Gutierrez, C., C. Hurtado, and A. Vaisman. *Temporal RDF*. 2005.
98. Rizzolo, F., et al., *Modeling concept evolution: A historical perspective*. 2009. p. 331-345.
99. Bykau, S., et al., *On Modeling and Querying Concept Evolution.* Journal on Data Semantics, 2012. **1**(1): p. 31-55.
100. Liao, H.-C. and C.-C. Tu, *A RDF and OWL-Based Temporal Context Reasoning Model for Smart Home.* Information Technology Journal, 2007. **6**: p. 1130-1138.
101. Gergatsoulis, M. and P. Lilis, *Multidimensional RDF*. 2005. p. 1188-1205.
102. Hartig, O. *Foundations of RDhÆ and SPARQnÆ (An Alternative Approach to Statement-Level Metadata in RDF)*. in *AMW*. 2017.
103. Dividino, R., et al., *Querying for Provenance, Trust, Uncertainty and other Meta Knowledge in RDF 1.* Web Semantics: Science, Services and Agents on the World Wide Web, 2009. **7**: p. 204-219.
104. Giunti, M., et al., *Representing n-ary relations in the Semantic Web.* Logic Journal of the IGPL, 2021. **29**(4): p. 697-717.
105. Batsakis, S. and E.G.M. Petrakis. *Representing Temporal Knowledge in the Semantic Web: The Extended 4D Fluents Approach*. in *Combinations of Intelligent Methods and Applications*. 2011. Berlin, Heidelberg: Springer Berlin Heidelberg.
106. Batsakis, S. and E.G.M. Petrakis. *SOWL: A Framework for Handling Spatio-temporal Information in OWL 2.0*. in *RuleML Europe*. 2011.
107. Milea, V., F. Frasincar, and U. Kaymak, *tOWL: A Temporal Web Ontology Language.* IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2012. **42**(1): p. 268-281.
108. Carroll, J., et al., *Named Graphs, Provenance and Trust.* Proceedings of the 14th international conference on World Wide Web WWW 05, 2004. **14**: p. 613-622.
109. Hernández, D., A. Hogan, and M. Krötzsch. *Reifying RDF: What works well with wikidata?* 2015.
110. Gutierrez, C., C. Hurtado, and A. Vaisman. *Temporal RDF*. in *The Semantic Web: Research and Applications*. 2005. Berlin, Heidelberg: Springer Berlin Heidelberg.
111. Li, H., *A New query method for the temporal RDF Model RDFMT Based on SPARQL*. 2021, Association for Computing Machinery: Chongqing, China. p. 1-6.

112. Zhang, F., et al., *Temporal Data Representation and Querying Based on RDF.* IEEE Access, 2019. **7**: p. 85000-85023.
113. Vinh, N., B. Olivier, and S. Amit, *Don't like RDF reification? making statements about statements using singleton property*. 2014, Association for Computing Machinery: Seoul, Korea. p. 759–770.
114. Keberle, N. *Temporal Classes and OWL*. in *OWLED*. 2009.
115. Consortium, W.W.W., *The RDF data cube vocabulary.* 2014.
116. Compton, M., et al., *The SSN ontology of the W3C semantic sensor network incubator group.* Journal of Web Semantics, 2012. **17**: p. 25-32.
117. Jacobsen, A., et al., *FAIR principles: interpretations and implementation considerations*. 2020, MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info …. p. 10-29.
118. Gangemi, A. and V. Presutti, *Ontology Design Patterns*. 2009. p. 221-243.
119. *R2RML: RDB to RDF Mapping Language*. 2012 [cited 2025 10/04/2025].
120. *RDF 1.1 Concepts and Abstract Syntax*. 2014 Last Access 13/04/2024]; Available from: https://www.w3.org/TR/rdf11-concepts/.
121. Gomez-Perez, A., O. Corcho-Garcia, and M. Fernandez-Lopez, *Ontological Engineering*. 2003: Springer-Verlag.
122. Zaveri, A., et al., *Quality assessment for linked data: A survey.* Semantic Web, 2016. **7**(1): p. 63-93.
123. Bellini, P., P. Nesi, and G. Pantaleo. *Benchmarking RDF Stores for Smart City Services*. 2015.
124. Zhang, L., et al., *BSBM+: Extending BSBM to Evaluate Annotated RDF Features on Graph Databases*. 2017. p. 129-143.
125. Gao, L., et al. *Stream WatDiv: A streaming RDF benchmark*. in *Proceedings of the International Workshop on Semantic Big Data*. 2018.
126. Schmidt, M., et al. *SP^2Bench: A SPARQL Performance Benchmark*. 2009.
127. Heling, L. and M. Acosta. *Federated SPARQL Query Processing over Heterogeneous Linked Data Fragments*. in *WWW 2022 - Proceedings of the ACM Web Conference 2022*. 2022.
128. Abuoda, G., et al. *Transforming RDF-star to Property Graphs: A Preliminary Analysis of Transformation Approaches*. 2022.
129. Fu, G., et al., *Exposing Provenance Metadata Using Different RDF Models.* 2015.
130. Frey, J., et al., *Evaluation of metadata representations in RDF stores.* Semantic Web, 2019. **10**(2): p. 205-229.
131. Berander, P. and P. Jönsson, *A goal question metric based approach for efficient measurement framework definition*, in *Proceedings of the 2006 ACM/IEEE international symposium on Empirical software engineering*. 2006, Association for Computing Machinery: Rio de Janeiro, Brazil. p. 316–325.
132. Braun, V. and V. Clarke, *Thematic analysis*, in *APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological.* 2012, American Psychological Association: Washington, DC, US. p. 57-71.
133. Jia, Z., et al., *TempQuestions: A Benchmark for Temporal Question Answering*, in *Companion Proceedings of the The Web Conference 2018*. 2018, International World Wide Web Conferences Steering Committee: Lyon, France. p. 1057–1062.
134. Jia, Z., et al., *Complex Temporal Question Answering on Knowledge Graphs*, in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. 2021, Association for Computing Machinery. p. 792–802.
135. Brooke, J., *SUS: A quick and dirty usability scale.* Usability Eval. Ind., 1995. **189**.
136. Sauro, J. and J.R. Lewis, *Quantifying the user experience: Practical statistics for user research*. 2016: Morgan Kaufmann.

137. Lewis, J.R. and J. Sauro. *The factor structure of the system usability scale*. in *Human Centered Design: First International Conference, HCD 2009, Held as Part of HCI International 2009, San Diego, CA, USA, July 19-24, 2009 Proceedings 1*. 2009. Springer.
138. Chorostowska-wynimko, J., *The Importance of Effective Registries in Pulmonary Diseases and How to Optimize Their Output.* 대한결핵및호흡기학회 추계학술발표초록집, 2021. **129**: p. 92-92.
139. Feltelius, N., et al., *Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.* Upsala journal of medical sciences, 2017. **122**(2): p. 136-147.
140. Ruseckaite, R., et al., *Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.* Patient related outcome measures, 2019. **10**: p. 355-372.
141. Stey, A.M., et al., *Clinical registries and quality measurement in surgery: a systematic review.* Surgery, 2015. **157**(2): p. 381-395.
142. Levay, C., *Policies to foster quality improvement registries: lessons from the Swedish case.* Journal of internal medicine, 2016. **279**(2): p. 160-172.
143. Kodra, Y., et al., *Recommendations for improving the quality of rare disease registries.* International journal of environmental research and public health, 2018. **15**(8): p. 1644.
144. Ogilvy, M. and J. Kollias, *Operating principles for running a clinical quality registry: are they feasible?* ANZ journal of surgery, 2012. **82**(11): p. 832-837.
145. Molina-Ortiz, E.I., A.C. Vega, and N.S. Calman, *Patient registries in primary care: essential element for quality improvement.* The Mount Sinai journal of medicine, New York, 2012. **79**(4): p. 475-480.
146. Health, A.I.o. and Welfare, *An AIHW framework for assessing data sources for population health monitoring: working paper*. 2014, AIHW: Canberra.
147. Long, J., J. Richards, and C. Seko, *The Canadian Institute for Health Information Data Quality Framework, Version 1: A Meta-Evaluation and Future Directions*. 2001. 370-383.
148. Yadav, P. and V. Samala. *Benchmarking over a Semantic Repository*. 2010.
149. Chen, L., S. Gao, and K. Anyanwu, *Efficiently evaluating skyline queries on RDF databases*. 2011. p. 123-138.
150. Minack, E., W. Siberski, and W. Nejdl. *Benchmarking fulltext search performance of RDF stores*. 2009.
151. Harbi, R., et al., *Evaluating SPARQL queries on massive RDF datasets.* Proc. VLDB Endow., 2015. **8**(12): p. 1848–1851.
152. Christina, E. and F. Burkhard, *An Approach to an efficient Evaluation of SPARQL Queries in Databases*. 2022, Association for Computing Machinery: Linz, Austria. p. 375–384.
153. Sen, S., et al., *RDFM: An alternative approach for representing, storing, and maintaining meta-knowledge in web of data.* Expert Systems with Applications, 2021. **179**: p. 115043.
154. Chawla, T., et al., *Storage, partitioning, indexing and retrieval in big rdf frameworks: A survey.* Computer Science Review, 2020. **38**.
155. Ermolayev, V., et al., *Ontologies of Time: Review and Trends.* International Journal of Computer Science and Applications, 2014. **11**: p. 57–115.
156. Kim, J., et al. *Query Performance Evaluation of OWL Storage Model*. 2011.
157. Kilintzis, V., N. Beredimas, and I. Chouvarda. *Evaluation of the performance of open-source RDBMS and triplestores for storing medical data over a web service*. 2014.
158. Haase, P., T. Mathäß, and M. Ziller, *An evaluation of approaches to federated query processing over linked data*, in *Proceedings of the 6th International Conference on Semantic Systems*. 2010, Association for Computing Machinery: Graz, Austria. p. Article 5.
159. Saleem, M., et al. *How representative is a SPARQL benchmark? An analysis of RDF triplestore benchmarks*. 2019.

Appendixes

A Thematic analysis – data source quality assessment framework

Literature reference	Extracted item from literature	Assigned Sub Theme1	Assigned Sub Theme2	Assigned Sub Theme3
Chorostowska-Wynimko et al (2019). The importance of effective registries in pulmonary diseases and how to optimise their output.[138]	Quality of structure	Governance	Finance	Organisation
Chorostowska-Wynimko et al (2019). The importance of effective registries in pulmonary diseases and how to optimise their output.	Quality of Data	Monitoring	Data Quality	
Chorostowska-Wynimko et al (2019). The importance of effective registries in pulmonary diseases and how to optimise their output.	Quality of Process	Monitoring	Ethics	
Feltelius et al (2017). Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.[139]	Establishing a management structure	Governance		

Literature reference	Extracted item from literature	Assigned Sub Theme1	Assigned Sub Theme2	Assigned Sub Theme3
Feltelius et al (2017). Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.	Agreeing a mandatory data subset	Governance		
Feltelius et al (2017). Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.	Preventing and monitoring incoherent entries	Monitoring		
Feltelius et al (2017). Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.	Motivating those submitting data	Training		
Feltelius et al (2017). Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.	Triangulation and data linkage of registry data to external data sources	Linkage		
Feltelius et al (2017). Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.	Data completeness in terms of patient population (as denominator)	Data Quality		
Feltelius et al (2017). Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.	Relevance of the data for answering the question	Data Quality		

Literature reference	Extracted item from literature	Assigned Sub Theme1	Assigned Sub Theme2	Assigned Sub Theme3
Feltelius et al (2017). Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.	Data granularity	Data Quality		
Feltelius et al (2017). Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.	Independence of the registry	Characteristics		
Feltelius et al (2017). Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.	Publications with data made from the registry	Report		
Feltelius et al (2017). Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.	Aspects of data protection.	Ethics		
Feltelius et al (2017). Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.	Organisational aspects	Organisation		
Feltelius et al (2017). Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.	Patient selection and coverage	Characteristics		

Literature reference	Extracted item from literature	Assigned Sub Theme1	Assigned Sub Theme2	Assigned Sub Theme3
Feltelius et al (2017). Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.	Data recorded	Data Management		
Feltelius et al (2017). Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.	Quality control procedures	Data quality	Monitoring	
Feltelius et al (2017). Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.	Governance	Governance		
Feltelius et al (2017). Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.	Reporting	Report		
Feltelius et al (2017). Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.	Data management and quality control	Data Quality	Data management	
Feltelius et al (2017). Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.	Ethical aspects	Ethics		

Literature reference	Extracted item from literature	Assigned Sub Theme1	Assigned Sub Theme2	Assigned Sub Theme3
Feltelius et al (2017). Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.	Legal, organisational, and financial aspects	Governance	Organisation	Finance
Feltelius et al (2017). Utility of registries for post-marketing evaluation of medicines. A survey of Swedish health care quality registries from a regulatory perspective.	Communication and reporting.	Report		
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.[140]	Purpose	Rationale		
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Stakeholders	Rationale		
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Eligibility	Characteristics		

Literature reference	Extracted item from literature	Assigned Sub Theme1	Assigned Sub Theme2	Assigned Sub Theme3
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Population	Characteristics		
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Sample size	Characteristics		
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Patient consent	Ethics		
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Patient involvement	Characteristics		
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	New/Existing	Characteristics		

Literature reference	Extracted item from literature	Assigned Sub Theme1	Assigned Sub Theme2	Assigned Sub Theme3
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Generic/specific	Characteristics		
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Timing & Frequency	Training		
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Mode & Method	Characteristics		
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Follow-Up	Characteristics		
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Entry & Quality check	Data Quality		

Literature reference	Extracted item from literature	Assigned Sub Theme1	Assigned Sub Theme2	Assigned Sub Theme3
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Storage	Data Management		
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Completeness	Data Quality		
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Descriptive analysis	Analysis		
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Inferential analysis	Analysis		
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Risk adjustment	Risk		

Literature reference	Extracted item from literature	Assigned Sub Theme1	Assigned Sub Theme2	Assigned Sub Theme3
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Benchmarking	Analysis		
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Time series	Analysis		
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Real-time analysis	Analysis		
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Dissemination	Report		
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Access to data	Access		

Literature reference	Extracted item from literature	Assigned Sub Theme1	Assigned Sub Theme2	Assigned Sub Theme3
Ruseckaite et al (2019). Developing a Preliminary Conceptual Framework for Guidelines on Inclusion of Patient Reported-Outcome Measures (PROMs) in Clinical Quality Registries.	Time & Frequency	Time		
Stey et al (2015). Clinical registries and quality measurement in surgery: a systematic review[141]	Standardise follow-up protocols; external validation and standardised training for abstractors; multiple imputation can be used for reliable quality comparison in the setting of missing data	Training	Monitoring	Standardisation
Stey et al (2015). Clinical registries and quality measurement in surgery: a systematic review	Shortened time until validation; cross-referencing and matching with National Death Index; implementing center-	Monitoring		

Literature reference	Extracted item from literature	Assigned Sub Theme1	Assigned Sub Theme2	Assigned Sub Theme3
	specific monitoring of missing data			
Stey et al (2015). Clinical registries and quality measurement in surgery: a systematic review	Reliability adjustment; standard definitions, adherence to a follow-up period for outcomes, and use of audits ; composite outcome measures vary more among hospitals than simple risk-adjusted rates; increase the number of cases sampled per institution or reliability adjustment; independent surgical nurse reviewer ; sharing cost with private sector	Standardisation	Monitoring	Data Quality

Literature reference	Extracted item from literature	Assigned Sub Theme1	Assigned Sub Theme2	Assigned Sub Theme3
Levay et al (2016). Policies to foster quality improvement registries: lessons from the Swedish case.[142]	Multi-level government funding of new and established registries	Finance		
Levay et al (2016). Policies to foster quality improvement registries: lessons from the Swedish case.	Focus on enabling quality improvement	Quality Improvement		
Levay et al (2016). Policies to foster quality improvement registries: lessons from the Swedish case.	Soft regulation and professional self-governance	Governance		
Levay et al (2016). Policies to foster quality improvement registries: lessons from the Swedish case.	Favourable patient data regulation	ethics		
Levay et al (2016). Policies to foster quality improvement registries: lessons from the Swedish case.	Clinical research	Research		
Levay et al (2016). Policies to foster quality improvement registries: lessons from the Swedish case.	Learning and collaboration between registries	Learning		
Levay et al (2016). Policies to foster quality improvement registries: lessons from the Swedish case.	Learning between registries	Learning		
Levay et al (2016). Policies to foster quality improvement registries: lessons from the Swedish case.	Automated data capture	Characteristics		

Literature reference	Extracted item from literature	Assigned Sub Theme1	Assigned Sub Theme2	Assigned Sub Theme3
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries[143]	definition and classification of registry	Definitions and Classifications		
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries	Governance	Governance		
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries	Data Elements, Case Report Form, Standardisations	Standardisation	Data management	Report
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries	IT Infrastructure Complying with FAIR Principles	Infrastructure		
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries	Data Quality	Data Quality		
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries	Quality Information	Data Quality		
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries	Documentation	Documentation		
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries	Training	Training		

Literature reference	Extracted item from literature	Assigned Sub Theme1	Assigned Sub Theme2	Assigned Sub Theme3
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries	Data Quality Audit	Monitoring		
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries	Definition of a Registry	Definitions and Classifications		
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries	Registry Classification	Definitions and Classifications		
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries	Defining Objectives	Rationale		
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries	Identifying Key Stakeholders Including the Particular Engagement of Patient Representatives	Organisation		
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries	Building the Team	Organisation		
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries	Ethical, Legal and Societal Issue (ELSI) and Privacy	Ethics	Governance	

Literature reference	Extracted item from literature	Assigned Sub Theme1	Assigned Sub Theme2	Assigned Sub Theme3
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries	Sustainability	Governance		
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries	Definition and Classification	Definitions and Classifications		
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries	Definition of Data Elements	Documentation		
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries	Establishment of Case Report Form (CRF) Questions	Report		
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries	Standardisations	Standardisation		
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries	Infrastructure Selection and Implementation	Infrastructure		
Kodra et al (2018). Recommendations for Improving the Quality of Rare Disease Registries	FAIR Data Principles	Standardisation		
Ogilvy et al (2012). Operating principles for running a clinical quality registry: are they feasible?[144]	Data collection	Data Management		
Ogilvy et al (2012). Operating principles for running a clinical quality registry: are they feasible?	Data elements	Data Management		

Literature reference	Extracted item from literature	Assigned Sub Theme1	Assigned Sub Theme2	Assigned Sub Theme3
Ogilvy et al (2012). Operating principles for running a clinical quality registry: are they feasible?	Risk adjustment	Risk		
Ogilvy et al (2012). Operating principles for running a clinical quality registry: are they feasible?	Data security	ethics		
Ogilvy et al (2012). Operating principles for running a clinical quality registry: are they feasible?	Data quality	Data Quality		
Ogilvy et al (2012). Operating principles for running a clinical quality registry: are they feasible?	Governance	Governance		
Ogilvy et al (2012). Operating principles for running a clinical quality registry: are they feasible?	Custodianship	ethics		
Ogilvy et al (2012). Operating principles for running a clinical quality registry: are they feasible?	Ethics/privacy	Ethics		
Ogilvy et al (2012). Operating principles for running a clinical quality registry: are they feasible?	Outputs	Report	Analysis	
Ogilvy et al (2012). Operating principles for running a clinical quality registry: are they feasible?	Resource	Finance	Infrastructure	

Literature reference	Extracted item from literature	Assigned Sub Theme1	Assigned Sub Theme2	Assigned Sub Theme3
Molina-Ortiz et al (2012). Patient registries in primary care: essential element for quality improvement.[145]	Identifies patients with diagnosis of interest Captures data elements electronically Offers real-time availability Is searchable Is a Web-based system linked to evidence-based guidelines Provides feedback to providers and healthcare team Generates reminder letters	Feedback	Characteristics	
An AIHW framework for assessing data sources for population health monitoring-Australian Institute of Health and Welfare (2014) [146]	Data source information	Characteristics		
An AIHW framework for assessing data sources for population health monitoring-Australian Institute of Health and Welfare (2014)	monitoring information area	Monitoring		

Literature reference	Extracted item from literature	Assigned Sub Theme1	Assigned Sub Theme2	Assigned Sub Theme3
An AIHW framework for assessing data sources for population health monitoring-Australian Institute of Health and Welfare (2014)	Data Quality	Data Quality		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017) [147]	Release and use	Time		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Value	Characteristics		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Adaptability	Characteristics		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Coverage	Characteristics		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Item availability	Characteristics		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Lineage	Characteristics		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Capture and collection	Characteristics		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Measurement error	Characteristics		

Literature reference	Extracted item from literature	Assigned Sub Theme1	Assigned Sub Theme2	Assigned Sub Theme3
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Processing	Characteristics		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Concordance	Characteristics		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Standardisation	Standardisation		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Linkage	Linkage		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Jurisdictional comparability	Characteristics		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Historical comparability	Characteristics		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Release timeliness	Time		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Release punctuality	Time		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Submission timeliness	Time		

Literature reference	Extracted item from literature	Assigned Sub Theme1	Assigned Sub Theme2	Assigned Sub Theme3
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Processing timeliness	Time		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Accessibility	Characteristics		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Interpretability	Characteristics		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Data source content	Data Management		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Data source purpose	Rationale		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Population of interest	Characteristics		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Data source type	Definitions and Classifications		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Maturity	Characteristics		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Data input file formats	Characteristics		

Literature reference	Extracted item from literature	Assigned Sub Theme1	Assigned Sub Theme2	Assigned Sub Theme3
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Submission	Characteristics		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Data provider types	Characteristics		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Relevance	Data Quality		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Accuracy and reliability	Data Quality		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Comparability and coherence	Data Quality		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Timeliness and punctuality	Data Quality		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	Accessibility and clarity	Data Quality		
CIHI's Data Source Assessment Tool- Canadian Institute for Health Information (2017)	General characteristics	Characteristics		

B Final papers in the literature review on temporal data modelling in healthcare

Title	Authors	Published Year
Abstraction on clinical data sequences: an object-oriented data model and a query language based on the event calculus	Combi, C.; Chittaro, L.	1999
Merging multimedia presentations and semistructured temporal data: a graph-based model and its application to clinical information	Combi, C.; Oliboni, B.; Rossato, R.	2005
Applying temporal joins to clinical databases	O'Connor, M. J.; Tu, S. W.; Musen, M. A.	1999
Spatio-temporal semantic representation of Cardiac MRI in heart attack patients	D. N. F. Awang Iskandar; H. Ujir	2015
A foundational model of time for heterogeneous clinical databases	Das, A. K.; Musen, M. A.	1997
Temporal representation of clinical algorithms using expert-system and database tools	Rucker, D. W.; Maron, D. J.; Shortliffe, E. H.	1990
Managing and displaying different time granularities of clinical information	Combi, C.; Pinciroli, F.; Musazzi, G.; Ponti, C.	1994
Modeling past, current, and future time in medical databases	Kouramajian, V.; Fowler, J.	1994
Coping efficiently with now-relative medical data	Stantic, B.; Terenziani, P.; Sattar, A.	2008
A Graph Database Approach for Temporal Modeling of Disease Progression	H. Memarzadeh; N. Ghadiri; S. P. Zarmehr	2018
A temporal model for Clinical Data Analytics language	L. Safari; J. D. Patrick	2013
CNTRO 2.0: A Harmonized Semantic Web Ontology for Temporal Relation Inferencing in Clinical Narratives	Tao, C.; Solbrig, H. R.; Chute, C. G.	2011
CNTRO: A Semantic Web Ontology for Temporal Relation Inferencing in Clinical Narratives	Tao, C.; Wei, W. Q.; Solbrig, H. R.; Savova, G.; Chute, C. G.	2010
Querying temporal clinical databases with different time granularities: the GCH-OSQL language	Combi, C.; Pinciroli, F.; Cavallaro, M.; Cucchi, G.	1995

Representing and querying now-relative relational medical data	Anselma, L.; Piovesan, L.; Stantic, B.; Terenziani, P.	2018
Exploiting bitemporal schema versions for managing an historical medical data warehouse: a case study	M. -. Serna-Encinas; M. Adiba	2005
Applying object-oriented technologies in modeling and querying temporally oriented clinical databases dealing with temporal granularity and indeterminacy	C. Combi; G. Cucchi; F. Pinciroli	1997
GCH-OSQL: a temporally-oriented object-oriented query language based on a three-valued logic	C. Combi; G. Cucchi	1997
Representation of Time-Relevant Common Data Elements in the Cancer Data Standards Repository: Statistical Evaluation of an Ontological Approach	Chen, H. W.; Du, J.; Song, H. Y.; Liu, X.; Jiang, G.; Tao, C.	2018
Temporal clinical data modeling and implementation for PTCA patients in an OODBMS environment	C. Combi; F. Pinciroli; G. Pozzi	1994

C Thematic Analysis –Metrics for evaluating temporal data modelling approaches in KGs

No	Article	Source	Metric	Polished metric	Sub theme	Parent theme
1.	Don't like RDF reification? Making statements about statements using singleton property [93]	S3	query execution time	Query response time	Performance	Querying

No	Article	Source	Metric	Polished metric	Sub theme	Parent theme
2.	Don't like RDF reification? Making statements about statements using singleton property	S3	Query length	Query length	Performance	Querying
3.	On Modeling and Querying Concept Evolution [99]	S3	running time	Query response time	Performance	Querying
4.	Benchmarking over a Semantic Repository [148]	S1	Execution time	Query response time	Performance	Querying
5.	Efficiently evaluating skyline queries on RDF databases [149]	S1	Execution time	Query response time	Performance	Querying
6.	Benchmarking fulltext search performance of RDF stores [150]	S1	Query evaluation time	Query response time	Performance	Querying
7.	Evaluating SPARQL queries on massive RDF datasets [151]	S1	Query response time	Query response time	Performance	Querying
8.	Evaluating SPARQL queries on massive RDF datasets [151]	S1	Data load time	Data load time	Performance	Querying

No	Article	Source	Metric	Polished metric	Sub theme	Parent theme
9.	An Approach to an efficient Evaluation of SPARQL Queries in Databases [152]	S1	Execution time	Query response time	Performance	Querying
10.	State-of-art approaches for meta-knowledge assertion in the web of data [17]	S2	Number of nodes	Number of nodes	Complexity	Modelling
11.	State-of-art approaches for meta-knowledge assertion in the web of data	S2	Information overhead (the number of statements)	Number of Statements	Complexity	Modelling
12.	State-of-art approaches for meta-knowledge assertion in the web of data	S2	Statement count	Number of Statements	Complexity	Modelling
13.	State-of-art approaches for meta-knowledge assertion in the web of data	S2	Redundant resource	Resource redundancy	Complexity	Modelling
14.	State-of-art approaches for meta-knowledge assertion in the web of data	S2	Storage space	Storage space	Performance	Uplifting

No	Article	Source	Metric	Polished metric	Sub theme	Parent theme
15.	State-of-art approaches for meta-knowledge assertion in the web of data	S2	Query length	Query length	Performance	Querying
16.	State-of-art approaches for meta-knowledge assertion in the web of data	S2	Query response time	Query response time	Performance	Querying
17.	Evaluation of meta data representations in RDF stores [130]	S2	Storage cost	Storage space	Performance	Uplifting
18.	Evaluation of meta data representations in RDF stores	S2	Mixed (metadata and data) query execution time	Query response time	Performance	Querying
19.	Evaluation of meta data representations in RDF stores	S2	Complexity	Complexity	Complexity	Modelling
20.	Evaluation of meta data representations in RDF stores	S2	Bulk load	Data load time	Performance	Querying

No	Article	Source	Metric	Polished metric	Sub theme	Parent theme
21.	Evaluation of meta data representations in RDF stores	S2	SPARQL conformance	W3C Query language support	Implementation support and standardisation	Querying
22.	Benchmarking RDF Metadata Representations: Reification, Singleton Property and RDF [15]	S1	Query execution time	Query response time	Performance	Querying
23.	Benchmarking RDF Metadata Representations: Reification, Singleton Property and RDF	S1	Complexity	Complexity	Complexity	Modelling
24.	Benchmarking RDF Metadata Representations: Reification, Singleton Property and RDF	S1	Storage size	Storage space	Performance	Uplifting
25.	Benchmarking RDF Metadata Representations: Reification, Singleton Property and RDF	S1	Adherence to standards	Adherence to w3c	Implementation support and standardisation	Querying/uplifting/modelling

No	Article	Source	Metric	Polished metric	Sub theme	Parent theme
26.	Benchmarking RDF Metadata Representations: Reification, Singleton Property and RDF	S1	Support by existing implementations	Implementation support	Implementation support and standardisation	Querying/uplifting/modelling
27.	RDF*M*: An alternative approach for representing, storing, and maintaining meta-knowledge in web of data [153]	S3	Additional triple generation	Additional triple generation	Complexity	Modelling
28.	RDF*M*: An alternative approach for representing, storing, and maintaining meta-knowledge in web of data	S3	Generation of redundant resources	Resource redundancy	Complexity	Modelling
29.	RDF*M*: An alternative approach for representing, storing, and maintaining meta-knowledge in web of data	S3	Query length requirement to execute a particular task	Query length	Performance	Querying
30.	RDF*M*: An alternative approach for representing,	S3	Number of triples	Number of triples	Complexity	Modelling

No	Article	Source	Metric	Polished metric	Sub theme	Parent theme
	storing, and maintaining meta-knowledge in web of data					
31.	RDF*M*: An alternative approach for representing, storing, and maintaining meta-knowledge in web of data	S3	Number of Statements	Number of Statements	Complexity	Modelling
32.	RDF*M*: An alternative approach for representing, storing, and maintaining meta-knowledge in web of data	S3	Resource redundancy	Resource redundancy	Complexity	Modelling
33.	RDF*M*: An alternative approach for representing, storing, and maintaining meta-knowledge in web of data	S3	Storage space requirement	Storage space	Performance	Uplifting
34.	RDF*M*: An alternative approach for representing, storing, and maintaining	S3	Query response time	Query response time	Performance	Querying

No	Article	Source	Metric	Polished metric	Sub theme	Parent theme
	meta-knowledge in web of data					
35.	Benchmarking RDF Stores for Smart City Services [123]	S1	data load time	Data load time in triple store	Performance	Querying
36.	Benchmarking RDF Stores for Smart City Services	S1	query execution time	Query response time	Performance	Querying
37.	Storage, partitioning, indexing and retrieval in Big RDF frameworks: A survey [154]	S1	data load time	Data load time	Performance	Querying
38.	Storage, partitioning, indexing and retrieval in Big RDF frameworks: A survey	S1	Data storage size	Storage space	Performance	Uplifting
39.	Storage, partitioning, indexing and retrieval in Big RDF frameworks: A survey	S1	query execution time	Query response time	Performance	Querying

No	Article	Source	Metric	Polished metric	Sub theme	Parent theme
40.	Ontologies of time: Review and trends [155]	S1	W3C Compliance	Adherence to w3c	Implementation support and standardisation	Querying/uplifting/modelling
41.	Ontologies of time: Review and trends	S1	Existing triple stores/implemented custom solution	Triple store support	Implementation support and standardisation	Querying
42.	Ontologies of time: Review and trends	S1	Interaction with a known query language	Query language support	Implementation support and standardisation	Querying
43.	Reifying RDF; What works well with Wikidata? [109]	S1	Query response time	Query response time	Performance	Querying
44.	Query Performance Evaluation of OWL Storage Model [156]	S1	Query response time	Query response time	Performance	Querying
45.	Evaluation of the performance of open-source RDBMS and triplestores for storing medical data over a web service [157]	S1	Query response time	Query response time	Performance	Querying

No	Article	Source	Metric	Polished metric	Sub theme	Parent theme
46.	An evaluation of approaches to federated query processing over linked data [158]	S1	Load time	Data load time	Performance	Querying
47.	An evaluation of approaches to federated query processing over linked data	S1	Query time	Query response time	Performance	Querying
48.	How Representative Is a SPARQL Benchmark? An Analysis of RDF Triplestore Benchmarks [159]	S1	Load Time	Data load time	Performance	Querying
49.	How Representative Is a SPARQL Benchmark? An Analysis of RDF Triplestore Benchmarks	S1	Storage Space	Storage space	Performance	Uplifting

D Survey for evaluating TEDME-KG metrics Framework.

You are being invited to participate in a study on evaluating a metrics framework which is used for evaluating temporal data modelling approaches in RDF-based Knowledge Graphs (KGs). This study is part of a research project that is supported by the ADAPT Research Centre, School of Computer Science and Statistics, Trinity College Dublin.

Procedure: There are three sections in this survey which are outlined below:

1. Assessment of the appropriateness of the metrics framework: In this section we want to have your opinion about the appropriateness of the framework for evaluating temporal data modelling approaches. We are using a set of criteria for assessing the appropriateness including relevant, unambiguous, direct, operational, understandable, and comprehensive. You will be asked to choose a score from 1 (Strongly disagree) to 5 (Strongly agree) for each statement based on the metrics framework.
2. Asking for suggestions: In this section we want to know your suggestions regarding the metrics framework. You will be asked to answer three questions.
3. Assessment of general usability and potential usefulness: In this section we want to have your opinion about the general usability and potential usefulness of the framework. You will be asked to choose a score from 1 (Strongly disagree) to 5 (Strongly agree) for each statement based on the metrics framework.

The detail of the metrics framework will be provided to you before answering the questions.

Participation in this study is completely **voluntary**. The survey takes approximately **30 minutes** to complete. The survey is **anonymous,** and your data will **not be identifiable.** The data will be stored using the IT services called MyZone Google Drive which complies with GDPR rules. The lead researcher (Sepideh Hooshafza) and the collaborator (Prof. Gaye Stephens) will be the only people with access to the data until its publication in an open data repository. None of your personal details will be recorded and you are free to stop and leave the experiment at any point if you so choose.

Data analysis: We will perform a quantitative analysis of the ratings using statistical summaries, reporting aggregated results. Qualitative analysis will be conducted manually by the lead researcher.

Publication: The result of this work will be published in relevant academic conferences and journals, as well as the PhD thesis of the lead researcher (Sepideh Hooshafza) at Trinity College Dublin.

Conflict of interest: The lead researcher and collaborator will not participate in the survey.

If you have any questions, please contact me: Sepideh Hooshafza, PhD Student (hooshafs@tcd.ie)

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 refuse to answer any question and that I may withdraw at any time without penalty.

● I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights.

● I understand that my participation is fully anonymous and that no personal details about me will be recorded.

● I understand that if I or anyone in my family has a history of epilepsy then I am proceeding at my own risk.

● I understand that personal information about me, including the transfer of this personal information about me outside of the EU, will be protected in accordance with the General Data Protection Regulation.

I consent to participate in this study, and consent to the data processing necessary to enable my participation and to achieve the research goals of this study. Yes/No

What is your current role in the field of RDF/Knowledge Graphs? *

PhD Researcher

Postdoctoral Researcher

How many years of experience do you have in RDF/Knowledge Graphs?*

1 < years <=2

2 < years <=5

5 < years

Metrics Framework Overview

Temporal data modelling in KGs refers to the practice of adding time to a triple. Different approaches of modelling temporal data in KGs exist including Singleton Property, Standard Reification, RDF*, Temporal RDF.

There is a lack of systematic approach for evaluating temporal data modelling approaches in KGs.

To address this gap, a metrics framework has been developed by the researchers of this study for evaluating temporal data modelling approaches in KGs. The metrics framework was developed from a thorough review of existing literature and through brainstorming sessions conducted by researchers during practical experiments.

This framework is organised into three main metrics categories:

1. Modelling
2. Uplifting
3. Querying

The framework further divided into three metrics subcategories:

1. Implementation support and standardisation
2. Complexity
3. Performance

The framework includes a total of 16 metrics.

For each metrics category a goal (Objective) has been defined and for each metrics subcategory, a corresponding list of questions has been created to facilitate evaluation.
You can find the metrics framework below (Table 1) or for better resolution, please download it at this link: Click Here

Section 1: Assessment of the appropriateness of the metrics framework

Please review the metrics and objectives (Table 1 or at this link) and answer the question below:

"Based on the metrics framework, to what extent do you agree or disagree with each statement?"

Please choose a score from 1 to 5.

- Relevant: The metrics framework is related to the fulfilment of objective.
- Unambiguous: The metrics framework is presented in a clear way that is obvious what is being measured.
- Direct: The metrics framework addresses objectives clearly.
- Operational: The metrics framework is practical and feasible.
- Understandable: The metrics framework is presented in a way that is easy to interpret.
- Comprehensive: The metrics framework thoroughly covers all the defined objectives.

If you have any comments about your response, please provide them in the space below:

Section 2: Asking for suggestion.

In this section we would like to know your suggestions regarding the metrics framework. You will be invited to answer three questions. You can find the metrics framework at this link: Click Here

1. Do you have any suggestion for adding a new goal and/or question in any of the metric categories or metric subcategories or changing the existing ones? Yes/No

Please provide more detail on your response.

2. Do you have any suggestion for adding a new metric, metrics category or metrics subcategory or changing the existing ones? Yes/No

Please provide more detail on your response.

3. Do you have any suggestion for grouping or labelling the metrics, metrics categories or metrics subcategories? Yes/No

Please provide more detail on your response.

Section3: Assessment of general usability and potential usefulness of the framework

In this section, we would like to gather your opinions on the general usability and potential usefulness of the framework.

You are invited to read a scenario, perform a few tasks, and then to respond to the 10 statements below. Please find the scenario, and metrics framework here: Click Here

Please read the scenario in section 1 and perform tasks in section 2.

Section 1: Scenario

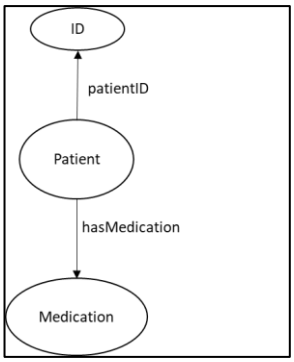
We have a dataset of 600 patients using different medications in different times. The data is stored in a relational database. The dataset contains data for patients who used a specific medication in a time interval (start date and stop date). Sample data for one patient is presented in Table 1.

Table 29-Sample data for a patient using medication in a time interval

Patient ID	Medication	Start date	Stop date
320	Prednisolone	05/06/2022	03/04/2024

We are going to model this dataset using knowledge graphs and answer a sample question: *"What is the duration of using Prednisolone for a patient with ID=320?"*

The ontology based on the dataset does not include medication time. (Figure 29)



We used two different approaches of temporal data modelling in KGs including "Singleton Property" and "Standard Reification" to add medication time to the ontology.

Approach 1: Singleton Property[7]

[7] More information on singleton property approach is in this link: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4350149/

The ontology including the medication time represented using "Singleton Property" is presented in Figure 2.

Figure 29- Ontology without including medication time.

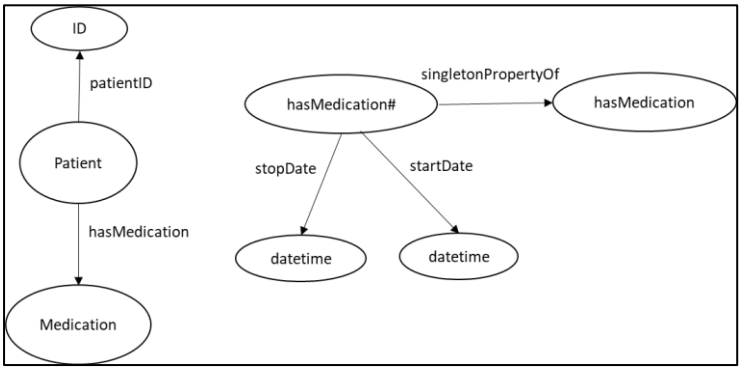


Figure 30- Modelling medication time using "Singleton Property" approach.

Approach 2: Standard Reification[8]

The ontology including the medication time represented using "Standard Reification" is presented in Figure 3.

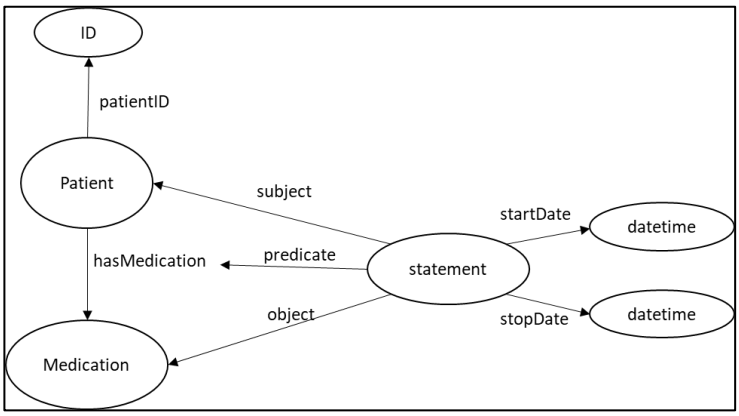


Figure 31- Modelling medication time using "Standard Reification".

Based on Approach 1 and Approach 2, we then uplifted data using R2RML engine (R2RML-F) and generated RDF data. The RDF data was imported into a triple store (GraphDB) and the query was designed.

[8] More information on standard reification approach is in this link: https://iccl.inf.tu-dresden.de/web/Inproceedings3037/en

We would like to evaluate approach 1 and approach 2 so we propose a metrics framework to evaluate temporal data modelling approaches in KGs. (See the metrics framework in the Folder)

Section 2: Tasks

You are an expert working in RDF knowledge graphs, you are going to evaluate two approaches discussed in the previous section. The below table presents the results of applying each metric in the proposed metrics framework to approach 1 and 2. Please do the following tasks to test the metrics framework:

Task 1: Evaluate approach 1 and 2 based on the complexity in modelling category

Please review the below table and see if you can evaluate the two approaches based on the results of the metrics in the modelling category.

Task 2: Evaluate approach 1 and 2 based on performance in querying

Please review the below table and see if you can evaluate the two approaches based on the results of the metrics in the performance in querying category.

Task 3: Evaluate approach 1 and 2 based on the implementation and standardisation in uplifting

Please review the below table and see if you can evaluate the two approaches based on the results of the metrics in the implementation support and standardisation in uplifting category.

Table 30- Results of applying metrics framework to approach 1 and 2

Metrics Category	Metrics Subcategory	Metric	Approach 1: Singleton Property	Approach 2: Standard Reification
Modelling	Implementation support and Standardisation	1. Ontology editor support	No	No
		2. Ontology adherence to W3C	No	No

Metrics Category	Metrics Subcategory	Metric	Approach 1: Singleton Property	Approach 2: Standard Reification
	Complexity	3. Number of nodes	38054 nodes	11636 nodes
		4. Number of triples	22904 triples	43348 triples
		5. Additional triple generation	13544 triples	27375 triples
		6. Resource redundancy	21137 triples and nodes	34479 triples and nodes
Uplifting	Implementation support and Standardisation	7. Uplift engine support	Yes	Yes
		8. Uplift adherence to W3C	Yes	Yes
	Complexity	9. Number of triple maps	4 triple maps	5 triple maps
		10. Number of term maps	14 term maps	17 term maps
	Performance	11. RDF Data storage space	2966 KB	5125 KB
Querying	Implementation support and Standardisation	12. W3C Query language support	Yes	Yes
		13. Triple store support	Yes	Yes
	Performance	14. Data load time	2000 ms	2000 ms
		15. Query length	6 lines	8 lines
		16. Query response time	2000 ms	4000 ms

Once you have completed the tasks, please return to the survey, and respond to the 10 statements.

To what extent do you agree or disagree with each statement? Please choose a score from 1 to 5 after completing the scenario above.

- I think that I would like to use such a metrics framework frequently.
- I found the metrics framework unnecessarily complex.
- I thought the metrics framework was easy to use.
- I think that I would need the support of a technical person to be able to use this metrics framework.
- I found the various parts of this metrics framework were well integrated.
- I thought there was too much inconsistency in this metrics framework.
- I would imagine that most Semantic Web researchers would learn this metrics framework very quickly.
- I found the metrics framework very cumbersome to use.
- I felt very confident using the metrics framework.
- I needed to learn a lot of things before I could get going with this metrics framework.

E Python code for calculating SUS score.

```python
import numpy as np

# Provided responses
responses = np.array([
    [5, 1, 5, 1, 5, 1, 5, 1, 5, 1],
    [4, 1, 4, 3, 5, 1, 5, 2, 4, 2],
    [3, 2, 4, 3, 4, 2, 4, 2, 3, 2],
    [5, 1, 5, 1, 5, 1, 5, 1, 4, 1],
    [5, 1, 4, 1, 5, 1, 5, 1, 5, 1],
    [5, 1, 5, 2, 4, 2, 4, 1, 4, 2],
    [4, 2, 4, 2, 4, 2, 4, 2, 4, 2],
    [5, 2, 5, 1, 5, 2, 4, 2, 3, 2],
    [5, 1, 5, 2, 4, 2, 4, 1, 4, 2],
    [4, 1, 4, 3, 5, 1, 5, 2, 4, 2]
])

# Calculate the score contribution
# Odd indices (0-based) are positive: 0, 2, 4, 6, 8
# Even indices (0-based) are negative: 1, 3, 5, 7, 9

positive_indices = [0, 2, 4, 6, 8]
negative_indices = [1, 3, 5, 7, 9]

positive_contributions = responses[:, positive_indices] - 1
negative_contributions = 5 - responses[:, negative_indices]

# Sum all contributions and multiply by 2.5 to get the final SUS score
total_contributions = np.sum(positive_contributions, axis=1) +
np.sum(negative_contributions, axis=1)
sus_scores = total_contributions * 2.5

# Calculate average SUS score
average_sus_score = np.mean(sus_scores)
print(average_sus_score)
```