

Digital Humanities and AI in Women's History: The VOICES Project (Ireland)¹

Lucy McKenna, Bronagh Ann McShane, Jane Ohlmeyer, Declan O'Sullivan, and Daniel Patterson

Summary

The first part of the chapter offers a brief introduction to the 'digital windfall' of Irish historical data (Part I) while part two discusses the innovative technologies and methods that the VOICES project will harness, including a Knowledge Graph and AI informed approaches. Two detailed case studies, that reflect 'work in progress' on records from the Irish Court of Chancery (Part III) and testamentary records (Part IV) are explored in the second half of the chapter. We reflect on the possibilities that machine learning affords historians and the processes we have been developing to avail of these opportunities. We argue that the transparent and ethical use of innovative technologies, grounded in the Semantic Web, Linked Open Data, and AI, can enable a new paradigm of historical research, which allows us to recover in a holistic and integrated way the marginalised voices and lived experiences of women in early modern Ireland.

Keywords: Ireland, women, gender, 1641 Depositions, wills, Court of Chancery, early modern, digital humanities, Knowledge Graph, Linked Data, Artificial Intelligence.

On 2 October 1672, Elizabeth Hammond, a 'spinster' living in Dublin city, made her last will and testament which she declared orally in front of several witnesses. In it, she made several bequests of money, amounting to the considerable sum of £740, along with the remainder of her estate, to family members and friends. In addition, she donated her 'little gold cupp' to her cousin, Mary Trevor. Her short will, which amounts to no more than 200 words, highlights the importance of her domestic network, especially her female relatives – her aunt, cousin, and two

sisters (who acted as her executors) – and her servant, Ann. These were the individuals who mattered most to Elizabeth. As well as revealing her cherished relationships, social status, and material wealth, the will offers a unique opportunity to hear the direct speech of a woman who otherwise left little trace in the historical record.²

The VOICES project (2023-2028) aims to recover the everyday experiences of non-elite women, like Elizabeth Hammond, who lived in early modern Ireland but who have largely been ignored by scholars more interested in privileging the stories of men of power and influence.³ Two ambitious research questions underpin VOICES. First, what role did women play in a society undergoing profound economic, political, religious, and cultural transformation? Second, what were their experiences of recurring social upheaval, bloody civil war, and extreme trauma, especially sexual violence? VOICES will recover these female life stories during times of peace and war (*c.*1550-*c.*1700) by drawing on a remarkable ‘digital windfall’ of diverse historical data which is now accessible alongside evolving digital methods that can make sense of and place these data into productive dialogue. In other words, VOICES merges empirical historical analysis of qualitative and quantitative data and storytelling with techniques pioneered by computer scientists to interrogate and visualise historical knowledge through linked data.

This chapter focuses on women in early modern Ireland, but the methodologies and machine learning techniques employed here have broader applications, extending to the study of women across diverse time periods and regions. The central aim of the VOICES project - to recover the lived experiences of women in early modern Ireland - aligns with longstanding efforts by historians of women and gender to amplify marginalised voices across temporal and geographical boundaries.⁴ The advent of AI and machine learning tools can help to address gaps left by centuries of gendered archival and research practices which have privileged patriarchal narratives and male perspectives. By expediting the process of locating women in

voluminous archival records, these tools can enable historians to address their lack of visibility more efficiently, thereby fostering a more inclusive and nuanced understanding of history. But we must proceed with caution. While harnessing the potential of AI, it is essential to remain mindful of its limitations, particularly the risks of algorithmic bias that can reinforce misinterpretation or exclusion of women. Ensuring transparency, ethical data use, and interdisciplinary collaboration is critical.⁵

This chapter will offer a brief introduction to the ‘digital windfall’ of historical data (Part I) before moving to outline the innovative technologies and methods that the project will harness (Part II). These include a Knowledge Graph and AI informed approaches: both ‘traditional’ (e.g. Transkribus, a Handwritten Text Recognition (HTR) platform) and ‘generative’ (e.g. Chat GPT-4o).⁶ Two detailed case studies, reflecting ‘work in progress’ on records from the Irish Court of Chancery (Part III) and testamentary records (Part IV) are explored in the second half of the chapter. Given that the VOICES project is in its infancy, this chapter outlines some possibilities that machine learning affords historians and the processes we have been developing to avail of these opportunities. It is important to state at the outset that the use of Transkribus and Chat GPT-4o in our case studies constitutes an early experiment aimed at gauging the utility and/or challenges of these resources for the purposes of generating and proofing transcriptions of early modern manuscript documents. In the case of Transkribus, some of the major limitations identified were not in fact limitations of the technology but rather of the quality of the digital images themselves. In the case of ChatGPT-4o, which is of course highly contested, the model was deployed in a controlled manner that involved manual verification of the output (i.e. a human in the loop approach).⁷ As will be highlighted, the importance of careful ‘prompt engineering’ in achieving the desired outputs cannot be overstated, demonstrating the inherent value of the ‘human’ in this human-machine collaboration.

The ‘Digital Windfall’

An unprecedented amount of Irish historical data, previously inaccessible, is now available, albeit in a variety of digital formats (digitisation of these records began in the late 1990s and is ongoing). This ‘digital windfall’ falls into four bodies of historical evidence: first records produced as bespoke digital humanities projects; second, digitised records held by public institutions; third, material digitised as part of the Virtual Record Treasury of Ireland (VRTI) project; and finally, records digitised and reconstituted for VOICES that, we hope, will ultimately form part of the VRTI. The totality of the material listed below means that we are almost drowning in digital data, but it is our intention, over the course of the VOICES project, to work through it all, together with other records that will become available in digital format.

1. Digital Humanities Projects: the Statute Staple Records and the 1641 Depositions.

The Statute Staple Records, which register lending and borrowing between the 1590s and 1680s, were edited from manuscripts (held in the British Library (BL) and Dublin City Library and Archive (DCLA)) and published in 1998 as a Microsoft Access database in a CD-ROM.⁸ Unreadable in the original format, the data was extracted into Excel spreadsheets, where it could be manipulated and the entries that related to female debtors and creditors could be isolated. Of the total number of creditors identified, women accounted for 5.5 per cent. In the case of debtors, the figure was less than 5 per cent. The Statute Staple data was exported from the Access database in CSV format to be parsed into bespoke person and event schemas developed for VRTI and VOICES. Overall, this was a straightforward process owing to the already well-organised data. The biggest challenge lay in identifying every woman, her associated loans, and every person associated with the same loans with absolute accuracy.

The 1641 Depositions comprise over 8,000 witness testimonies, examinations, and associated materials in which men and women of all backgrounds told of their experiences following the outbreak of violence on 22 October 1641.⁹ In 2010 the 1641 Depositions were published online, a process that took three researchers three years to transcribe the entire archive and to mark them up in TEI (Text Encoding Initiative). As the corpus was carefully curated manually, while not perfect, we are confident of its accuracy and it now serves as ‘the gold standard’ for the fully or semi-automated processes of research that we are undertaking as we identify the lived experiences of women. In all, 958 women deposed, often more than once.¹⁰ They were mostly Protestant, and many were first- or second-generation migrants predominately from England but also Wales and Scotland. In other words, roughly one eighth (or 12 per cent) of all deponents **were** female, which by early modern standards is a very high proportion. The number of single women 576 (of 958) or 60 per cent is also very high with ‘widows’ forming a majority (511 or 53 per cent of the total).¹¹ These female deponents came from every province and every county and from rural and urban areas (see Figure 1). Hundreds more women are named in the depositions given by men. The majority are ordinary, non-elite women from the ‘middling sort’, a group so often absent from history.

The process of curating the data for the 1641 Depositions was substantively like that undertaken for the Statute Staple, albeit much more involved. The Depositions required the addition of several data categories to the ‘person schema’ and the development of an extensive ‘deposition event schema’ to reflect satisfactorily the complexity and upheaval represented in the original documents. Unique data categories were needed to capture geographical movement, widowhood and remarriage, the loss of property, and ambiguous occupational data encoded using the system developed by the Cambridge Group for the History of Population.¹²

female. To date, both the PCC wills and the Funeral Entries have been largely untapped by historians of early modern Ireland.¹³

The Dublin Directory, prepared by DCLA, is a digital orphan that is no longer available online but the archivist kindly provided Excel spreadsheets of the cess (or tax) paid by the citizens of Dublin between 1647 and 1649; of the pipe water accounts for 1680, 1681, and 1694; of the city's principal inhabitants in 1684; and of women admitted to the city's franchise between 1564 and 1750. Though limited to Dublin, the detail about the lived experiences of women contained in these records is invaluable to the VOICES project.¹⁴

3. **The Virtual Record Treasury of Ireland** is a landmark initiative which aims to reconstitute Ireland's National Archives destroyed in 1922.¹⁵ The collaboration with VRTI affords the VOICES project access to exceptionally important historical data – inquisitions, wills, surveys, censuses, fiants, petitions, the Books of Survey and Distribution, the Lodge Collection, and other records dating from early modern period feared lost¹⁶ – together with digital copies of books published by the Irish Manuscripts Commission,¹⁷ the *Dictionary of Irish Biography*, with biographies of 11,000 people (roughly ten per cent of whom were elite women),¹⁸ and Ordinance Survey maps and place names.¹⁹ There is a considerable overlap in the AI based technologies being deployed and collaboration between the projects is ongoing.

Particularly valuable is the Lodge Collection, 20 volumes in the National Archives of Ireland compiled in the early eighteenth-century by the archivist and genealogist, John Lodge (1692–1774), which have just been fully digitised but never fully utilised. There is much in Lodge relating to women and especially to marriage settlements, dowers (a widow's share of her husband's estate), and jointures (land that was allocated to a woman at marriage).²⁰ Even more important are the Books of Survey and Distribution, the

equivalent of the Domesday Book (a great English land survey dating from 1086) for Ireland, which capture the revolution in land holding that the country experienced during these years.²¹ Drawn up in the wake of the restoration of Charles II and the victory of William III, in part for taxation purposes ('quit rent'), these contain a wealth of information. The names of land holders, including women, in 1641, in *c.*1670, and again in *c.*1704 appear in columns together with the number of plantation acres (profitable and non-profitable) and the county, barony and parish where the land was held. Some entries contain information on mortgages, jointures, and dowers. At least five per cent (*c.*295 out of *c.*9,200) of the people mentioned as landholders in 1641 were women, although the final figure may be much higher.

4. Records digitised or reconstituted for VOICES.

We are digitising a substantial proportion of the Irish Chancery pleadings held at the National Archives of Ireland (NAI) which survived the 1922 fire. We have captured a corpus of 1,167 documents where women are listed as plaintiffs, defendants, or both (discussed in Part III below). Unlike Irish Chancery pleadings which are held in a single archive, testamentary records are highly fragmented, widely dispersed across multiple archives and in many cases no longer extant.²² In England, large numbers of original testamentary sources survive for the sixteenth and seventeenth centuries and have informed (and continue to inform) important large-scale social and cultural studies, but the same cannot be said for Ireland. Despite these challenges, we are attempting to reconstitute what has survived (discussed in Part IV below).²³

In short, the windfall from digitisation is exceptional but the resulting data remains unstructured, existing in 'digital silos' and in a variety of formats, from pdf files and CD

ROMs to datasets. How then do we transform our ‘digital windfall’ into knowledge about women that is interoperable across different formats and digital silos?

The Knowledge Graph and AI technologies

One of the founders of the Semantic Web has referred to the condition of most data on the Web as a:

‘data wilderness - full of valuable treasure, but overgrown and tangled. Even the valuable data that you can find can take any number of forms, adapted to its own part of the wilderness’.²⁴

This description maps neatly onto the problem presented by the tangle of digital data currently available for VOICES which is in a variety of digital formats: unstructured (e.g. 1641 Depositions); semi-structured (e.g. Statute Staple records); data linked to the VRTI Knowledge Graph; and digital images (e.g. testamentary records). The VOICES Knowledge Graph turns this ‘data wilderness’ into data that can be interrogated and visualised in meaningful and innovative ways.

A ‘Knowledge Graph’²⁵ (KG) can be considered a network of data, whose nodes and connections are described by an ontology. In simple terms, an ontology is a way to organise concepts in a formal manner, to ease understanding (for humans and computer applications) of how terms used in a domain are related. Guidelines and principles of how these ontologies and knowledge graphs can be represented in a standardised manner, such that they enable the publication of data and knowledge as FAIR (Findable, Accessible, Interoperable and Reusable), have been provided through Semantic Web and Linked Open Data initiatives.²⁶

The VOICES KG is grounded in the CIDOC Conceptual Reference Model (CIDOC-CRM) ontology²⁷. CIDOC-CRM is a high-level, event-centric ontology used to represent relationships found in historical and cultural heritage data, enabling the structured

description of people, events, objects, and the contexts that connect them. VOICES extends CIDOC-CRM by introducing concepts specifically for modelling women, their roles in early modern Ireland and their experiences/activities during this period. These concepts form part of a series of controlled vocabularies that can be used to complete elements of the person and event schemas e.g. named titles, occupation, social status, event type, event roles, types of violence, women's roles. Where possible, pre-existing controlled vocabularies are being used and extended for VOICES (e.g. occupational data is being encoded using the system developed by the Cambridge Group for the History of Population).

Prior to data being added to the VOICES KG, it must first be curated according to a bespoke person schema which was initially developed for the VRTI KG and then extended for the VOICES project to capture more female centric information. The person schema and its editorial guidelines were developed in close collaboration with the historians. The schema contains 53 fields grouped into categories such as:

- Names and aliases (modernised and variant spellings)
- Biographical dates and places
- Familial relationships
- Occupations and social status
- Links to external identifiers and references (e.g., Wikidata, DIB)

The minimum requirements for an entry in the person schema is a last name, an address and dating information (i.e. floruit, date of birth, date of death). However, most entries for women contain much more detailed information. We capture a woman's maiden name, where known, as her headline name, and her married name(s) as an alias. Names are modernised but variant spellings are also captured. Data about male relatives, primarily husbands, is also captured using the person schema but these are typically stub entries that only meet the schema's minimum requirements.

Data on events and activities that women participated in and experienced is also being captured through a separate event schema. Events currently being modelled include depositions (from the 1641 Depositions), loans (from the Statute Staple records), probate (the Prob 11 Wills data) and legal proceedings (from the Irish Chancery records). The event schema captures, at minimum, the type of event in question, the dates associated with the event, the place in which an event took place, and the people/groups involved and their respective roles. More detailed information on events is captured by extending the schema for each event type. Once the person and event schemas for a particular dataset are complete, the data is then uplifted to the KG.

In the field of computer science, the use of KGs, the Semantic Web and Linked Open Data are increasingly proving their worth but have rarely been adapted for the challenge of the ‘messy’ or ‘dirty’ historical data that have survived from past centuries. As such, their use here will be highly innovative and advance the state of the art within both computer science and history. VOICES adopts novel methods by combining analysis of qualitative, quantitative, and cartographic data with techniques pioneered by computer scientists for the interrogation and visualisation of historical knowledge through KG representation and making the KG available as Linked Open Data. It does not require the digitisation of new archives but rather the linking of existing data and the creation of bespoke controlled vocabularies for women and violence that will feed into the VOICES KG. It gathers information from both subject matter experts (e.g. historians collecting information or social scientists working with numeric data) as well as automated techniques (e.g. named entity recognition in digitised documents).

Though VOICES has drawn inspiration from the VRTI project, it moves significantly beyond the current state of the art by focussing on the curation of historical data to generate new knowledge about the lived and wartime experiences of women. The VOICES KG will allow for the rigorous interrogation, analysis, visualisation of all data, and linking of that data

with datasets, some of which had been marked up in TEI, from other scholars, using a variety of existing tools (as the technologies underpinning the KG are based on open internet standards developed by the World Wide Web Consortium (W3C)). In other words, an AI processing pipeline will be created to allow for semantic classification of concepts in ontologies and linking of entities, which then allows for interrogation, visualization and analysis. The pipeline takes unstructured or semi-structured data, provides techniques for digital transcription processing, entity extraction, declarative semantic mapping, and tools for data representation, querying, and interlinking. Declarative semantic mapping is a technique that allows for the specification of how data can be taken from a dataset in one format and defined via a schema (e.g. spreadsheet with a user/project defined format) and transformed into another dataset format according to another schema (e.g. RDF with CIDOC-CRM standardised ontology), all the while preserving the semantics of the data transformed. Across the pipeline, the ontology-driven approach facilitates different disciplinary perspectives arising from the datasets. Its success depends on a deep exchange between the computer scientist and the historian whose modes of conceptualizing knowledge can be markedly different.

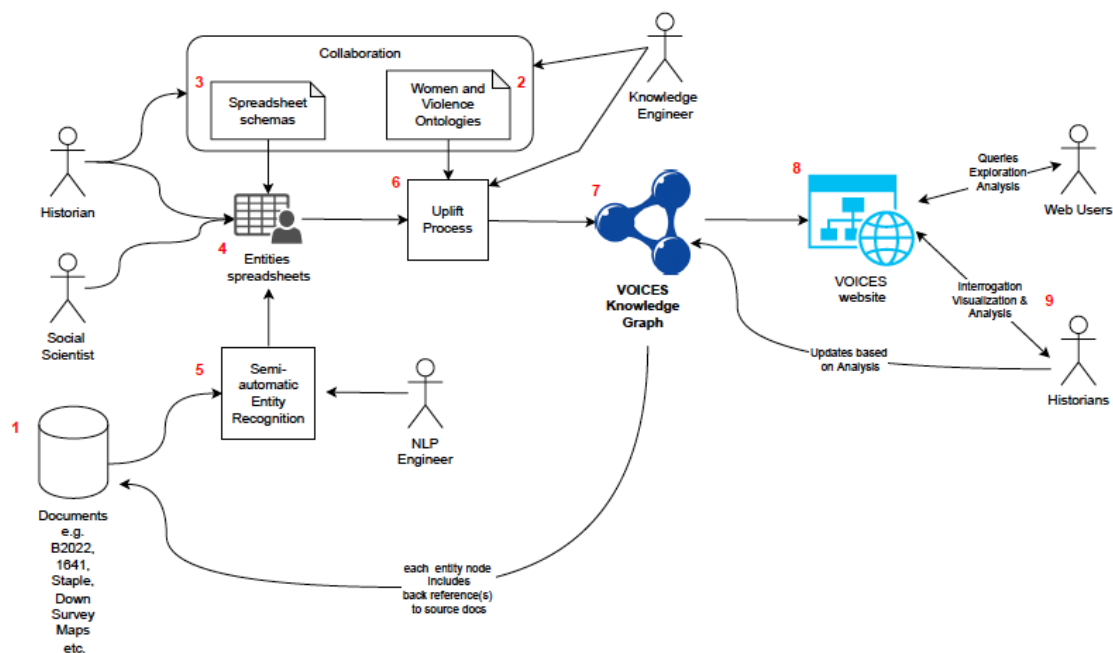


Figure 2: Diagram of the technical workflow (WP0)

As Figure 2 (above) shows VOICES does not require the digitisation of new archives but rather the linking of existing data (#1) and the creation of bespoke controlled vocabularies for women and violence (#2) that will feed into a VOICES KG (#7). However, VOICES has also been experimenting with the use of traditional AI tools (e.g. Transkribus) and emerging generative AI tools (e.g. ChatGPT-4o) to turn parts of existing archives into digitally accessible transcriptions. Since its inception in 2013, Transkribus has become the most popular user-facing platform for producing transcripts of historical texts across the cultural and heritage industries. Its impact on academic research, especially in the humanities domain, has been considerable (though as yet not fully quantifiable).²⁸ Countless humanities research projects focusing on a broad range of subjects from medieval to modern have harnessed the technology, reporting varying levels of accuracy and success depending on type of handwritten script in question, with early modern studies well represented.²⁹ In the case of early modern records, the platform is especially valuable in making archival material that was hitherto the preserve of a handful of specialists trained in palaeography, accessible to non-specialists. Parts III and IV below discuss two case studies which illustrate early experiments in using Transkribus and ChatGPT-4o to produce and proof automatically generated transcriptions.

The availability of automatically generated transcriptions allows for: a) immediate digital availability and analysis of historical information by historians, and b) further semi-automatic processing of the transcriptions by Natural Language Processing (NLP) tools (#5) to extract entities of interest within the KG. The KG gathers information from both subject matter experts (e.g. historians collecting information or social scientists working with numeric data, #4) as well as semi-automated techniques (e.g., Named Entity Recognition in digitised documents, #5). Named Entity Recognition (NER) processing is non-trivial for historical data, as the identity of the entity type that the word references needs to be identified through examination in the context of the rest of the text (e.g. the word ‘Baker’ in a sentence could

reference an entity of type 'occupation', or it could be surname, which would indicate that the word references an entity of type 'person'), and then identify and disambiguate the particular entity (e.g. perhaps it is referencing a surname of a person already identified in the KG, in which case a link to this document should be made, or maybe a new person has just been identified that needs to be added to the KG). In other words, for every noun encountered, the entity type needs to be determined, the entity disambiguated and possibly uniquely identified. The output of candidate entity matches needs curation by historians.

Information is added to the KG by the automatic uplift process (#6) which extracts entities from the person and event schema spreadsheets and creates a KG according to the ontology. The approach uses concepts, relations, and instances to represent entities and their connections in various sources. Put simply, the ontology captures the types of entities and how they relate to each other, for example, one should see the concept of 'Task', but not specific tasks (e.g. 'spinning') as these are captured using the bespoke controlled vocabularies. The historians populate schema spreadsheets (#4) centred around key concepts or entities (e.g. Gender, Name, Date of Birth, Address, Status, Nationality, Religion, Task, Land, etc). These spreadsheets are uplifted into the VOICES KG (#7), where the data is linked according to the ontologies. Editorial guidelines (or 'spreadsheet schemas', #3), compiled by historians and the knowledge engineer (an expert in the ontology and knowledge graph approach), on how to fill in these spreadsheets are crucial to manage the identification of entities (for the 'entities spreadsheets', #4) and for consistency of data input. We will ensure, in line with Open Research guidelines, that the process is transparent so that it can be verified, replicated, and used in collaborative platforms.

The VOICES KG when finalised will be made available remotely via the web for querying and analysis both to software applications and human users by adopting the W3C Linked Data approach (#8). The current Beta version of the VOICES KG brings together

material from five distinct archival collections—the 1641 Depositions, the Statute Staple, the Freewomen of Dublin, and a selection of women’s wills from TNA, PROB 11. Together, these collections generated 2,345 entries for Irish women from the early modern period for the VOICES KG. This marks a transformative expansion as only 54 early modern Irish women were previously represented in the VRTI KG making VOICES a major step toward restoring their presence in the historical record. A video demo³⁰ of the Beta VOICES KG explores the entries of five of these women via the VRTI Knowledge Graph website³¹ which itself is still under development, currently version 1.2.

Court of Chancery records

The records of the Irish Court of Chancery constitute arguably the most significant and ‘complete’ corpus of legal documents available to scholars of early modern Ireland. They form the largest single collection of records which survived the destruction of the Irish Public Record Office in 1922. In the last quarter of the twentieth century, the special significance and richness of Irish Chancery for early modernists was underlined in a series of publications which exhibited promising findings and invited further research.³² Chancery’s emerging status as a major source for Irish history seemed sure to be confirmed when, between 1999 and 2014, 5265 of the approximately 12,000 surviving pleadings were conserved by the NAI. Manuscripts which were previously brittle and covered in soot were rendered robust and readable.³³ Since conservation, however, Irish Chancery has received little attention from historians, remaining lamentably understudied and shrouded in obscurity.

The primary barrier to research progress for the Irish Chancery pleadings is their sheer scale and impenetrability.³⁴ Although there is some suggestion that the Irish pleadings were recovered from the wreckage of 1922 in their original ‘bundles’ – perhaps organised chronologically – it is difficult to know how closely the current listings resemble any pre-1922

structure. The bundles were quite literally plucked from the ashes in the weeks following the disaster, placed in brown paper parcels, and tied up with string ‘to be dealt with later’.³⁵ The result is an archive which – despite the labours of conservators and the historians who have worked on it – is difficult to use and understand. In 1972, Nicholls encountered ‘twenty-eight bundles into which the surviving Ancient Pleadings have been – haphazardly, except for the last two bundles of the series – gathered up’. Evidently some degree of reorganisation of the collection took place in the ensuing years; Mary O’Dowd’s article, written in 1999, refers to documents organised in boxes, albeit using ostensibly the same alphanumerical referencing system. For the avoidance of ambiguity, we refer to ‘series’ A, B, C, and so-on, in lieu of boxes or bundles.

As noted above, Nicholls alludes to the ‘final two’ series of Chancery pleadings as somehow less disorganised than the rest of the archive. Here he is referring to series AA and BB. Both consist almost exclusively of ‘answers’ – that is, the initial pleading supplied the defendant(s) to the plaintiff’s ‘bill of complaint’ which commenced a lawsuit at Chancery.³⁶ AA proceeds roughly chronologically from 1618-19 then 1622-27 with variously sized gaps of months. Interestingly, BB proceeds in almost perfect chronological order from March to December 1622, with concentrations of documents in May, July, and November - reflecting the terms of the legal year. Series BB is therefore the most likely to reflect accurately the structure of the original Chancery archive. Comparison with extant draft indexes, which were incomplete at the time of the 1922 fire, shows that the Answers in AA and BB were in the process of being indexed, and remain in their original order.³⁷

Interestingly, the vast majority of surviving pleadings in series A-Z do not appear in the indexes, and the fractured and incomplete nature of the remaining collection is most evident in this section of the archive. Most of the documents in series A-Z are undated bills of complaint, which make up approximately three quarters of the total collection. The remaining quarter

consists mostly of answers, along with some replications, rejoinders, demurrers, petitions, commissions, inventories, and other miscellaneous survivals.³⁸ Where the material is undated, internal evidence places it within the period 1570-1630.³⁹ A considerable body of additional material, including evidence which would have been presented to the court including depositions and exhibits, unfortunately does not survive to any considerable extent.⁴⁰

O'Dowd's pioneering article on Chancery as a court of redress for women in early modern Ireland sits alongside a much more developed historiography concerning gender and equity courts in early modern England.⁴¹ It is for this reason – along with the scale and relative scholarly neglect of the archive – that the VOICES project identified the Irish Chancery pleadings as a key target source. However, unlike most of the other sources which feed into VOICES, Chancery documents exist only in their original manuscript form. It was therefore necessary to digitise records which represented the lived experiences of women.

Our initial work therefore consisted of a scoping exercise to locate the women amongst the swathes of Chancery litigation, which like any other archive dating from the sixteenth and seventeenth centuries, bias heavily in favour of men. In this, we set out to build on the work of O'Dowd and Nicholls, both of whom had been able to carry out apparently unrestricted manual examinations of the original bundles/boxes of documents prior to conservation. During his work on Irish Chancery in the 1970s, Nicholls created an informal catalogue of all Chancery Pleadings in series A-Z and AA-BB. The list is a working document mostly consisting of reference numbers, inferred titles of the lawsuits,⁴² and occasional estimated date ranges for pleadings. Some pleadings are accompanied with a summary of the lawsuit, whilst others are fully or partially transcribed. Pleadings chosen for transcription seem to reflect Nicholls's own research interests, with the most fully described and transcribed examples relating to disputes over landholding, particularly those involving Gaelic Irish families.

The version of Nicholls's list available for our purposes was an untidy rendering (using Optical Character Recognition or OCR) of a document which was originally produced on a typewriter. It was therefore very difficult to use. Our first task was to read carefully the entire list and parse relevant cases into a spreadsheet to improve the accessibility and useability of the data. Initially, the sole criteria for inclusion on the list was *any* mention of a woman or girl anywhere in the title or in Nicholls's summary description/transcription of a pleading. This resulted in an initial list of 1,505 documents. A review of the list indicated that many of the pleadings which mentioned women but did not name them as litigants (in other words, as plaintiff(s), defendant(s), or both) often made only scanty reference to them. As such, to make the volume of digitisation more manageable and selective, we decided to revise the list to include only those documents where women were named as litigants, resulting in a shortlist of 1,212 documents. When we finally came to photograph the shortlisted items, a small number could not be located, or were determined not to involve women after all. Ultimately, we digitised a total of 1,167 Chancery pleadings.

It is worth noting here that the interventionist curation of such a list – and the resulting Knowledge Graph/digital 'archive' – is not ideal from the perspective of 'traditional' historical scholarship, since it imposes artificial limits for researchers who use it to study the collection. Women appear frequently in Chancery cases with nominally exclusively male litigants (to a greater extent than suggested by our 'initial list') and the representation of women in such contexts is also of interest to historians of women and gender.⁴³ However, digitisation of the entire collection would have been far beyond the scope of this project in its current iteration. We hope that this work will provide proof of concept for future work in the digitisation and Knowledge Graph spaces, radically improving the accessibility of archives including Irish Chancery.

An examination of our target list in conjunction with Nicholls's full catalogue of salved pleadings serves to confirm the findings presented in O'Dowd's 1999 article. For the corpus as a whole, female litigants appear in around 21% of cases. 12% of all Chancery plaintiffs are women, and 10% of defendants. Of the 1,167 pleadings which have been confirmed to involve women, 680 or 58 per cent involve female plaintiffs. 551 or 47 percent involved women as defendants, and approximately 8 per cent involved women on both sides. We have confirmed Mary O'Dowd's finding that women appeared as sole plaintiffs – without a man – 38% of the time. In the vast majority of cases, these women were widows. We calculated a slightly higher rate of women appearing as joint plaintiffs without the involvement of a man; 3.8% to O'Dowd's 1.7%. In most cases, these are widows and daughters, or orphan sisters. Our count also aligned with O'Dowd's in identifying that a majority – by our estimation 58% – of women plaintiffs appeared alongside men, usually their husbands. Further quantification of women's involvement in Chancery, their marital statuses, and other social and economic factors will be made possible once the entire corpus is digitised and uplifted to the VOICES Knowledge Graph.

With the assistance of staff at NAI, we photographed 1,167 items for mass automated transcription using Transkribus. In developing our approach to processing the Irish Chancery pleadings using Transkribus, we consulted with Emily Kadens, a legal scholar with a particular specialism in developing Transkribus models proficient at 'reading' sixteenth- and early seventeenth-century secretary hand in English legal documents. Professor Kadens has generously granted us access to her bespoke *Egerton.4* Text Recognition model, which has been trained on hundreds of thousands of words of English Chancery documents.⁴⁴

The first step with Transkribus was to use Professor Kadens's proprietary *Equity Baselines* model to identify the text regions and text lines in the Irish pleadings. Because the model was trained on very similar English Chancery pleadings, we have achieved a relatively

high degree of accuracy with the automated layout process. Text regions and lines were checked and, where necessary, corrected.

From here we proceeded to generate transcriptions using *Egerton.4*. The results varied from the incredibly accurate to the utterly incomprehensible. It should be noted that, almost without exception, cases where the Transkribus model produces unusable outputs results from the poor condition of the initial document or, occasionally, quality of image. The spectrum of quality is evident in the below figures.

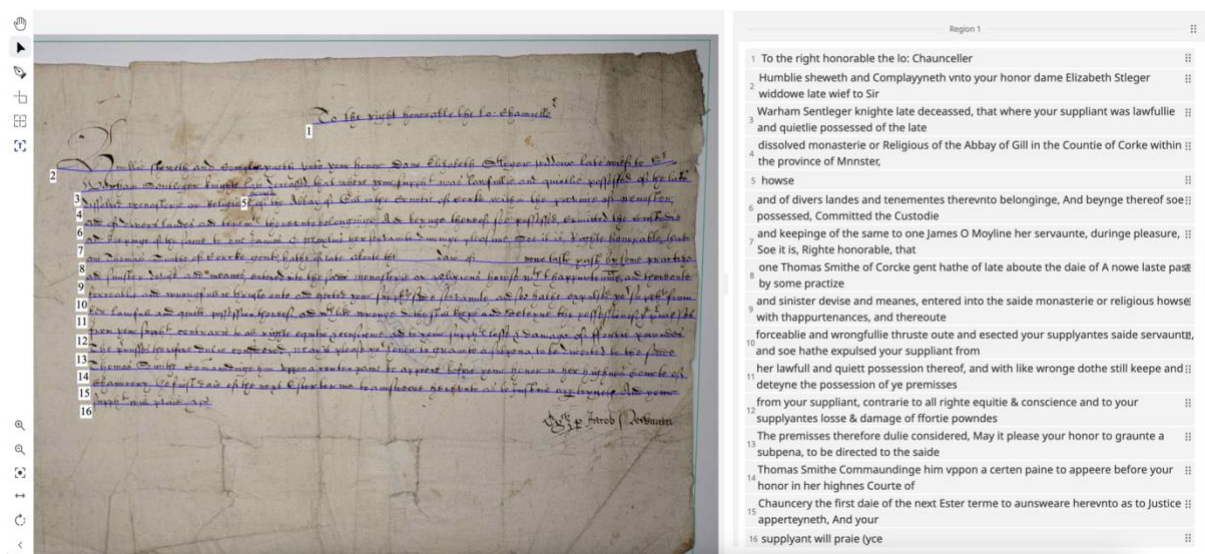


Figure 3: NAI, B/295, Dame Elizabeth St Leger widow of Sir Warham St Leger v. Thomas Smithe of Cork, undated

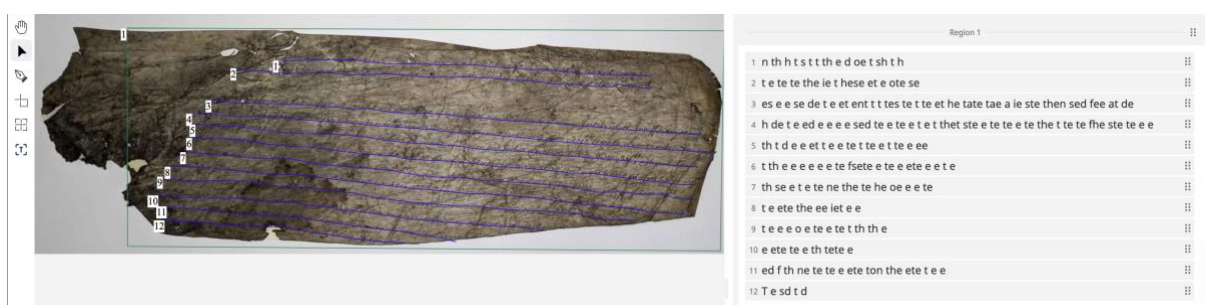


Figure 4: NAI, B/429, Fragment of Answer of Hugh Montgomery and Anne his wife to the bill of Sir Robert Maxwell, undated

The first (Figure 3) represents a very clear document and high-quality image, resulting in a near-perfect automated transcription before any manual checking or correction. The second (Figure 4) represents a document which was in very poor condition, with low contrast between

the tone of the vellum (which in this case was badly stained by dirt or soot) and the text itself. Most of the automated transcriptions sit somewhere between these two poles in terms of quality. The next piece of the puzzle lies in determining how we will tackle checking and, where necessary, correcting the remaining corpus of imperfect automatically generated text, whilst maintaining the VOICES project's emphasis on working at scale to capture as much data as possible. The transcriptions, though imperfect, allow a more straightforward 'entry point' into these relatively dense and complex documents, allowing researchers to quickly locate and check the names of litigants, geographical locations, and legal themes. Working between these transcriptions, the document images, and the data collated in Nicholls's catalogue, we have been able to begin compiling a schema of Chancery litigants for uplift to the VOICES Knowledge Graph. Although this work is still in progress, we estimate that the digitised documents will yield approximately 3,500 individual litigants, approximately 1,200 of whom are women.

The Chancery pleadings represent an unparalleled opportunity for the study of women in early modern Ireland. O'Dowd's work has already demonstrated the potential of these records, which afford fleeting glimpses at some elusive figures; 'native' Irish-speaking women, the 'poore mayde', the destitute daughter denied her inheritance, appearing *in forma pauperis*.⁴⁵ They contain hundreds of 'real-world' examples of disputes involving jointures, marriage settlements, portions, dowries, and wills, all of which formed the ever-changing social, legal, and economic backdrop to early modern Irish women's lives.⁴⁶ The VOICES project will identify, aggregate, and process data from this and over 1,000 other Chancery cases to enable research into Irish women's history in a way previously thought impossible.

Testamentary Records

Testamentary records for early modern Ireland can be described as the epitome of an archive ‘dispersed’.⁴⁷ The destruction of the Irish Public Records Office in 1922 resulted in the loss of virtually all original wills proved before the local consistorial courts and the Prerogative Court of Armagh. As a result, historians of early modern Ireland are compelled to rely largely on abstracts and copies of original wills made by nineteenth century antiquarians and genealogists (which present their own challenges, especially when it comes to reconstructing the lives of women). However, one important collection of *original* wills survives in TNA Kew: those proved in the Prerogative Court of Canterbury in London that were made by women and men (mostly of English background) who lived in Ireland but who had some property interests in England. While by no means representative of women in early modern Ireland as a whole (concerning, for the most part, one particular cohort of the island’s ethnically and religiously diverse population), the PCC wills are the single largest surviving collection of original wills made by women who lived in seventeenth century Ireland.⁴⁸ They are, therefore, of intrinsic value and constitute an instructive case study in the application of machine learning and AI techniques to early modern records made by and about women. As the case of Elizabeth Hammond cited at the beginning of this chapter demonstrates, the PCC wills present a unique opportunity to hear the ‘voices’, albeit in a mediated form, of women who otherwise rarely appear in the historical record.⁴⁹

Within the PCC collection at TNA, the PROB 11 series contains the will registers, that is copies of original wills entered into large volume (registers) by church court clerks when the wills were proved before the PCC. Where the original version of the will survives, it appears in a separate series known as PROB 10 (which have for the most part not been digitised).⁵⁰ Microfilms of the PROB 11 registered copies have been digitised and these ‘digital surrogates’ are available to download as pdf files on TNA’s website.⁵¹

As part of their efforts to reconstruct material lost during the destruction of the Irish Public Records Office, the VRTI project scraped data from TNA's PROB 11 listing extracting from this corpus 'Irish hits' (about 4,000 wills relating to Ireland, extracted from the 2,263 volumes of wills).⁵² This data has been incorporated into the VRTI's 'repository' layer for TNA and are now available to search on their website.⁵³ From these c.4,000 'Irish hits' on the VRTI website, all wills made between 1550 and 1720 were extracted, the dates that the VOICES project is focused on. This returned 572 'Irish' wills of which a total of 57 (or 10 per cent) are by women. Of these 57, the majority, 29, are by women who either lived in or had some sort of association with Dublin, but there is a good deal of representation beyond Dublin, with wills made by women living in all four provinces of Ireland included among the 57. The geographical association of the testatrices is derived from the will title, for example, 'Will of Ellen Poore of Waterford, Waterford', which presents its own methodological challenges in terms of identifying wills by women with links to Ireland.

In terms of chronological scope, the earliest will dates to 1623, when Magdaelene Ring, a young woman of just thirteen years old from Mocollop, County Waterford, made her last testament.⁵⁴ In terms of social status, the women are not surprisingly from the more well to do ranks of society, those who actually possessed enough material wealth and property to make bequests in the first place. However, only a minority are from what we might deem to be the very upper echelons of society, the marker of which was the holding of a title: there are just five dames, three countesses and one viscountess among the 57 women. Instead, most are from that 'middling sort', a cohort which became a defining feature of early modern society.⁵⁵ Intriguingly, the corpus also includes wills by two women who describe themselves as servants, indicative of the potential of the PCC collection for yielding insights into a subsection of society that so rarely appear in the archives. Not surprisingly, the majority of female will makers are widows, reflecting the legal apparatus of early modern society which restricted a

married woman's rights to hold and exchange property: a married woman required a husband's consent to dispose of personal property and, therefore, to make a will (this remained the case until 1882).⁵⁶ There were, of course, exceptions to this general rule as is demonstrated in the case of Susannah Feilding alias Booth of Dublin who in August 1700 made her will, explaining that, 'notwithstanding my coverture', a clause in her marriage settlement empowered her 'to dispose by will in writing any summe of money not exceeding five hundred pounds'.⁵⁷

Wills made by women with links to Ireland in the PCC collection were used as a case study to perform some initial experiments with Transkribus and generative AI (in this case ChatGPT-4o) in order to produce and proof automatic transcriptions that could then be subject to further in-depth analysis. In widening access to readable transcripts of manuscript sources, Transkribus presents exciting opportunities that can illuminate new research horizons and insights. But it is not a silver bullet. Early modern handwriting was deeply heterogenous: it varied from individual to individual and featured a preponderance of variant spelling coupled with the use of abbreviations, and capitalisation. Furthermore, there was considerable variation in handwriting style over the time period we designate as 'early modern'. Thus, as Martin Prell has recently cautioned, in the case of early modern handwritten sources at least, entirely error-free automatic recognition will likely never become a reality, and the interpreting editor will never be replaced by the machine.⁵⁸ While early modern scholars can rest easy in the knowledge that they will not be replaced by a machine any time soon, the potential of the Transkribus platform to offer a 'way in' to previously inaccessible material, is certainly something to be excited about. For a project such as VOICES, that seeks to recover the lives and experiences of women in early modern Ireland and to do so at scale, the possibilities are promising.

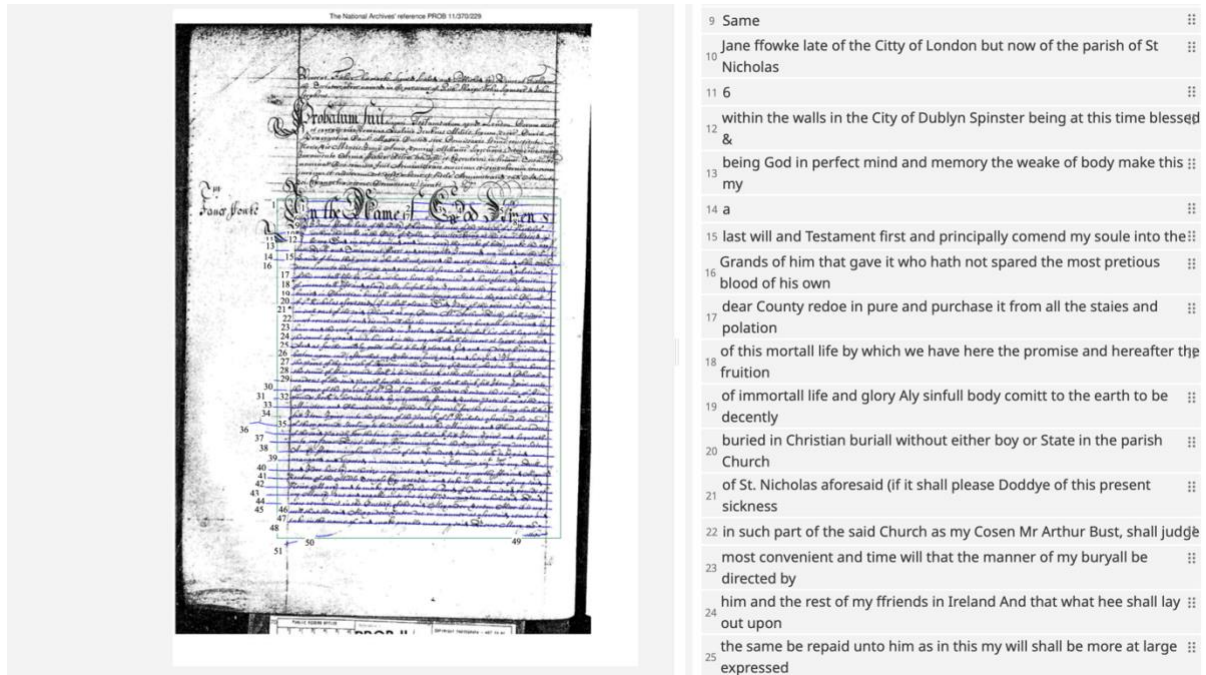


Figure 5: Will of Jane Fowke, spinster of Saint Nicholas Dublin, County Dublin, TNA, PROB 11/370/229

The first step in generating the automatic transcriptions involved uploading the digital images of the wills to Transkribus. From here, it was necessary to conduct an image layout analysis, which involves labelling text regions and lines and is a crucial preprocessing step that significantly enhances the expected character recognition rate. The images were then run through one of the platform’s publicly available ‘out-of-the-box’ text recognition models (in this case The Text Titan I, the platform’s ‘Super Model’).⁵⁹ Given the small sample size and the fact that this was an early experiment, we did not deem it time efficient to develop our own customised ‘wills model’, although there is currently one being developed by the ‘Material Culture of Wills’ project’, which we eagerly await.⁶⁰

There were two key challenges with this experiment. First was the quality of the digital images themselves; they are what is known as ‘noisy’ images i.e. the image quality is poor. The digital images downloaded from TNA’s website are black and white microfilm images which, as Harry Smith and Emily Vine have discussed, come with several inherent problems, even for manual transcription purposes. Their black and white rendering means that low-contrast

regions are hard to read while text ‘is also frequently obscured by folds or creases on the page’. In addition, ‘bleed through text from the other side of the parchment also occasionally renders the microfilms unreadable’.⁶¹ For these reasons, those who have transcribed PROB 11 wills by hand have pointed to the difficulty of reading marginalia and inter-lined text on the digitised microfilms, something which cannot be rectified other than by viewing the original volume (which as Smith and Vine have shown, is not always feasible or possible).⁶² These problems with the ‘digital surrogates’ were, not surprisingly, reflected in the accuracy of the Transkribus generated transcriptions and must be viewed as a deficiency of the source images rather than of the technology itself.

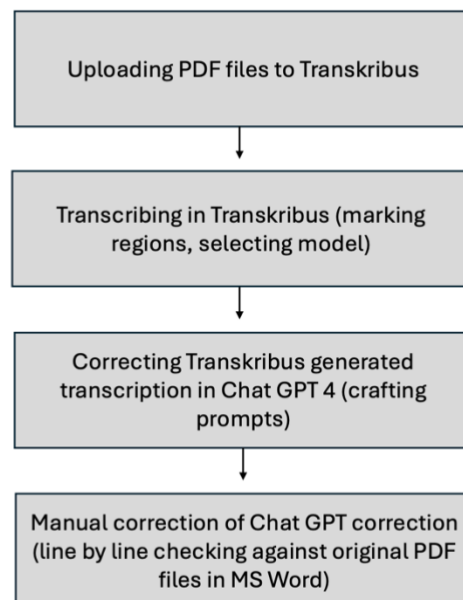


Figure 6: Experimental process for generating and proofing automatic transcriptions

A second challenge is that the publicly available Text Titan Super Model struggled to get to grips with the particular hand used in the registered copies of the wills; secretary’s hand. Given that secretary’s hand varies considerably even over the chronological span that the wills in this case study cover (1623-1720), this is not surprising. The identification of named entities within the text (especially names of people and places) was very haphazard and could vary

considerably even within a single text. For example, the name ‘Wilkinson’ is rendered ‘Wilomson’, ‘Wilimson’ and ‘Wiskinson’ in the same will.⁶³ This was amplified further in cases where the will is a translation into English from French, as is true in the case of Judith Jaquean (Aug. 1699) and Mary De St. Julien (Feb. 1707), both members of the French Huguenot migrant community living in Dublin at the turn of the eighteenth century.⁶⁴ In the case of placenames rendered in their Irish language form (which are rare but reveal something of the linguistic ‘ambidexterity’ of the female testatrixes in question), the model struggled.⁶⁵ For example, in her will of March 1631, Frances Parker described herself as being of ‘the town and Countie of Catherlaugh (the Irish form of ‘Carlow’). This is rendered as the ‘Countie of Catherlangly’ in the automated transcription.⁶⁶ Given that Titan is not trained on Irish language text, this inaccuracy is not surprising.⁶⁷ Undoubtedly better quality images would have mitigated some of the difficulties identified here. Ensuring accuracy of named entities and places is crucial when it comes to conducting subsequent analysis, for example, via Named Entity Recognition processing, which as discussed in Part II, is non-trivial and relies on accurate transcriptions.

The second step in this experiment involved using ChatGPT-4o to automatically proof and correct the Transkribus-generated transcriptions. The first step in the process involved the creation of appropriate prompts to enable the model to proof effectively and correct errors within the transcriptions in such a way that would not alter the text i.e. by changing or adding punctuation or modernising language/spelling. The importance of thoughtful prompt engineering in guiding the model’s output cannot be overstated. It is an iterative process and one that has, according to Quizi Guo, becomes ‘not merely a task’ but a ‘new art form in the digital age’. As Guo has contended not only is prompt engineering ‘about posing the right questions’, it is ‘also about initiating and sustaining a dialogue [and] reaching the desired outcome collaboratively’. When choreographed correctly, this ‘dance between user and model’

can, Guo argues, ‘lead to a symphony of insights’.⁶⁸ This experimental ‘dance’ with ChatGPT-4o began using the following very basic prompt:

- Prompt 1: ‘This is a Transkribus generated transcription of a seventeenth-century will. I want you to proof and correct it’.

Upon reviewing the output, it was immediately clear that the model had modernised the spelling and added punctuation which was not in the raw transcription. Clearly this was not an ideal output for the purposes of further analysis since it was not an accurate reflection of the original text. A further caveat was then added to the prompt:

- Prompt 2: ‘This is a Transkribus generated transcription of a seventeenth-century will. I want you to proof and correct it. Do not modernise the spelling or add punctuation’.

With this additional instruction, the desired output was improved. The model corrected many (though not all) of the errors in the Transkribus-generated transcripts. However, it was important to review the model’s output after each transcript was corrected to ensure consistency. For example, after completing the task for the first few transcriptions, the model slipped back into the modernising of spelling. This necessitated the reiteration of the importance of not modernising spelling by means of a third prompt:

- Prompt 3: ‘This is a Transkribus generated transcription of a seventeenth-century will. I want you to proof and correct it. Do not modernise the spelling or add punctuation. Please do not modernise the spelling’.

This process was carried out for each of the 57 will transcriptions, with a total of 44,522 words proofed and corrected by ChatGPT-4o. In each case, the model’s output was manually checked against the original image of the will, for verification and to finalise the transcription. The model performed surprisingly (though not consistently) well, especially when it came to correcting named entities in the script. For example, the name Parker is rendered as ‘Parker’, ‘Larker’ and ‘Parter’ in the automated transcription but the model correctly guessed that

‘Larker’ and ‘Parter’ should in fact be ‘Parker’. But the necessity to manually verify ChatGPT-4o’s output and the model’s lack of ‘explainability’ ultimately calls into question its utility for humanities research, especially for a project such as VOICES, which seeks to process data automatically, at scale, and in a transparent and ethical manner. Further exploratory work into the generation of customised and source specific prompts, in addition to the construction of ‘walled gardens’,⁶⁹ offer a way forward that may, in time, allow us to harness the potential of generative AI in a manner that does not compromise the historian’s quest for empirical knowledge.

Conclusion

A central purpose of the VOICES project is to demonstrate that the transparent and ethical use of innovative technologies, grounded in the Semantic Web, Linked Open Data, and AI, can transform unstructured historical data into knowledge about women that can be interrogated and visualised. This is not to say that such techniques constitute a panacea or are free from technical or ethical challenges. The controlled vocabularies on women and violence that form part of VOICES will be published openly (using the W3C Ontology Web Language format) to allow for interoperability with other research initiatives and for their reuse by others. We hope that by adopting a transparent approach, we will provide the proof of concept for scholars sceptical of these machine learning technologies, and a supranational methodological model for others who are hesitant to apply these approaches for lack of guidance on the technical aspects. Thus, VOICES will provide the tools, methods, and semantic data to make it work for a new generation of scholars. It will also enable a new paradigm of historical research, which allows us to recover in a holistic and integrated way the marginalised voices, lifecycles, and identities of women in early modern Ireland; to assess their contribution to the household, regional and national economies and their relationship to the land in an era of colonisation,

commercialisation, and expropriation; and to better understand the life stories of women such as Elizabeth Hammond and thousands of others.

Figure Captions

Figure 1. The geographic distribution of women who deposed in 1641.

Source: <https://voices.adaptcentre.ie/map.html>.

Figure 2. Diagram of the technical workflow (WP0)

Source: Voices

Figure 3. B/295, Dame Elizabeth St Leger widow of Sir Warham St Leger v. Thomas Smithe of Cork, undated.

Source: National Archives of Ireland

Figure 4: B/429, Fragment of Answer of Hugh Montgomery and Anne his wife to the bill of Sir Robert Maxwell, undated.

Source: National Archives of Ireland

Figure 5. Will of Jane Fowke, spinster of Saint Nicholas Dublin, County Dublin, TNA, PROB 11/370/229.

Source: The National Archives

Figure 6. Experimental process for generating and proofing automatic transcriptions.

Source: Voices

Notes

¹ This project has received funding from the European Research Council (ERC) under the European Union's Horizon Europe Research & Innovation programme under grant agreement No 101097003. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the ERC. Neither the European Union nor the granting authority can be held responsible for them. Jane Ohlmeyer is the Principal Investigator for VOICES, <https://voicesproject.ie/>, accessed 17 October 2024. This is a jointly authored article with each team member drafting particular sections: Jane Ohlmeyer (Part I), Declan O'Sullivan and Lucy McKenna (Part II), Daniel Patterson (Part III), and Bronagh Ann McShane (Part IV). Professor O'Sullivan and Dr McKenna are Computer Scientists involved in the Research Ireland Research Centre ADAPT (grant number 13/RC/2106-P2).

² Will of Elizabeth Hammond or Hamond, Spinster of Dublin, County Dublin, The National Archives, London [hereafter TNA], PROB 11/341/302.

³ Notable exceptions in Ireland are the pioneering works by Margaret MacCurtain, Mary O'Dowd, Clodagh Tait, Naomi McAreevey, and Marie-Louise Coolahan. For a recent account of the state of the art in women's and gender history in Ireland see, Frances Nolan and Bronagh Ann McShane, 'Introduction: A new agenda for women's and gender history in Ireland', special issue of *Irish Historical Studies*, 46:170 (2022), pp. 207-16 [<https://doi.org/10.1017/ihs.2022.26>].

⁴ See, for example, Joan W. Scott, 'Gender: a useful category of historical analysis', *The American Historical Review*, 91:5 (1986), pp. 1053-75, a foundational early work concerning the methodologies of gender history and the need to address the silences in historical narratives.

⁵ Jon Chun and Katherine Elkins, 'The crisis of artificial intelligence: A new digital humanities curriculum for human-centred AI', *International Journal of Humanities and Arts Computing*, 17:2 (2023), pp. 147-67 [<https://doi.org/10.3366/ijhac.2023.0310>].

⁶ Transkribus, a Handwritten Text Recognition (HTR) platform that generates machine processable text from images of historical manuscripts, developed by the READ-Coop, <https://www.transkribus.org>; and Generative AI, in this case ChatGPT-4o ('o' for 'omni') is the most recent successor of ChatGPT-4, the fourth in a series of GPT (Generative Pre-trained Transformers) natural language processing models developed by OpenAI, <https://openai.com>. ChatGPT-4o was released in May 2024 and was accessed via a paid subscription, allowing for significantly higher usage limits than the free version. The experiments conducted with ChatGPT-4o were conducted between May and August 2024.

⁷ For a thoughtful recent reflection on the use of AI and ChatGPT in history research see Moira Donovan, 'How AI is helping historians better understand our past', *MIT Technology Review*, April 2023. <https://www.technologyreview.com/2023/04/11/1071104/ai-helping-historians-analyze-past/>, accessed 18 October 2024.

⁸ Jane Ohlmeyer and Éamonn Ó Ciardha (eds.), *The Irish Statute Staple Books, 1596-1687* (Four Courts Press: Dublin, 1998).

⁹ <https://1641.tcd.ie>, accessed 18 October 2024.

¹⁰ In other publications the number of female deponents was recorded as 959 but one deponent – Blagram Swan of Tipperary (TCD, Ms 821, ff. 94r) – was in fact the male servant of Dorothy Bastard.

¹¹ Figures from Jane Ohlmeyer, *Making Empire. Ireland, Imperialism and the Early Modern World* (Oxford University Press: Oxford, 2023), chapter 3.

¹² E. A. Wrigley, 'The PST system of classifying occupations', *CAMPOP Preliminary Reports*, published online <<https://www.campop.geog.cam.ac.uk/research/projects/occupations/britain19c/papers/paper1.pdf>> accessed 19 October 2024.

¹³ <https://catalogue.nli.ie/Record/vtls000580170>, accessed 26 October.

¹⁴ Jane Ohlmeyer, 'The lived experiences of women in early modern Dublin: John T. Gilbert Commemorative Lecture, 29 May 2024' (DCLA: Dublin, forthcoming, 2025).

¹⁵ <https://www.virtualtreasury.ie/> accessed 18 October 2024. VRTI is creating a repository of manuscript and print digital images combined with transcriptions generated automatically using *Transkribus* Handwritten Text Recognition technology and born-digital TEI-XML.

¹⁶ Micheál Ó Siochrá, 'Lost and found: 'new' sources for early modern Irish history' in Marian Lyons, Terence Dooley and Salvador Ryan (eds.), *The historian as detective: uncovering Irish pasts* (Four Courts Press: Dublin, 2021), pp. 277-280.

¹⁷ <https://www.irishmanuscripts.ie>, accessed 18 October 2024.

¹⁸ <https://www.dib.ie/about>, accessed 18 October 2024.

¹⁹ <https://store.osi.ie/>, accessed 18 October 2024.

²⁰ The Lodge material is listed in the *55th Report of the Deputy Keeper of Public Records in Ireland* (1928), pp. 116–22. The deeds calendared by Lodge always include the names of both parties when the deed is held jointly by husband and wife.

²¹ The most complete set is in the National Archives in Dublin (the ‘quit rent’ set) and includes entries for every county in Ireland.

²² For a detailed discussion of testamentary sources for early modern Ireland see Clodagh Tait, ‘Writing the social and cultural history of Ireland, 1550-1660: wills as example and inspiration’ in Sarah Covington, Vincent P. Carey and Valerie McGowan-Doyle (eds.), *Early modern Ireland. New sources, methods and perspectives* (Routledge: Abingdon, 2019), pp. 27-48; Clodagh Tait, ‘“As legacie upon my soule”: The wills of the Irish Catholic community, c.1550-c.1660’ in Robert Armstrong and Tadhg Ó Hannracháin (eds.), *Communities in early modern Ireland* (Four Courts Press: Dublin, 2006), pp. 179-98.

²³ For a major large-scale project on wills in early modern England see, ‘The Material Culture of Wills, England 1540-1790’ project, led by Jane Whittle at the University of Exeter and funded by the Leverhulme Trust (<https://sites.exeter.ac.uk/materialcultureofwills/>)

²⁴ Dean Allemang, James Hendler, Fabien Gandon, *Semantic Web for the Working Ontologist: Effective Modeling for Linked Data, RDFS and OWL* (3rd edition, 2020).

²⁵ ‘A knowledge graph, also known as a semantic network, represents a network of real-world entities—i.e. objects, events, situations, or concepts—and illustrates the relationship between them. This information is usually stored in a graph database and visualized as a graph structure, prompting the term knowledge “graph”, <https://www.ibm.com/cloud/learn/knowledge-graph>, accessed 18 October 2024.

²⁶ Ontologies are commonly defined as ‘a [formal] explicit specification of a [shared] conceptualization’, Christophe Debruyne, Gary Munnely, Lynn Kilgallon, Declan O’Sullivan and Peter Crooks, ‘Creating a Knowledge Graph for Ireland’s Lost History: Knowledge Engineering and Curation in the Beyond 2022 Project’, *ACM Journal on Computing and Cultural Heritage* (2022), pp. 1-25.

²⁷ <https://cidoc-crm.org/>

²⁸ Joe Nockels, Paul Gooding, Sarah Ames and Melissa Terras, ‘Understanding the application of handwritten text recognition technology in heritage contexts: a systematic review of Transkribus in published research’, *Archival Science* (2022) 22, pp. 367–92.

²⁹ See for example, Martin Prell ‘Early modern letters as a challenge automated handwriting recognition. Experience and Results Report on the Transkribus Project’ (2018) [<https://doi.org/10.22032/dbt.34849>] and C. Annemieke Romein, Sara Veldhoen and Michel de Gruijter, ‘The datafication of early modern ordinances’, *DH Ben- elux, Leiden*, 2, (2020), [<http://dx.doi.org/10.17613/80sx-m116>]

³⁰ <https://youtu.be/P9XSB9HbfwM>

³¹ <https://kg.virtualtreasury.ie/>

³² The VOICES project is very grateful to NAI for generously facilitating and assisting the research outlined below. Kenneth Nicholls, ‘Some documents on Irish law and custom in the sixteenth century’ in *Analecta Hibernica*, 6 (1970), pp. 103–29; ‘A calendar of salved Chancery Pleadings concerning County Louth’, *Journal of the County Louth Archaeological and Historical Society*, 17 (4) (1972), pp. 250–60 and 18 (2) (1974), pp. 112–20; Mary O’Dowd, ‘Women and the Irish Chancery Court in the late sixteenth and early seventeenth centuries’, *Irish Historical Studies*, 31 (124) (1999), pp. 470–87; Jane Ohlmeyer, ‘Records of the Irish Court of Chancery: A preliminary report, 1627–1634’ in Desmond Greer and Norma Dawson (eds.), *Mysteries and Solutions in Irish Legal History* (Four Courts Press: Dublin, 2001), pp. 15–50. O’Dowd also draws on Chancery material in *A history of women in Ireland, 1500–1800* (Routledge: London, 2004), though the original research was carried out prior to conservation.

³³ Zoë Reid and Benjamin van de Wetering, ‘Conservation of the sixteenth- and seventeenth-century Chancery pleadings at the National Archives of Ireland’, *Journal of the Institute of Conservation*, 35 (2013), pp. 189–200.

³⁴ The same explanation has been offered for the relative neglect of English Chancery – a far larger though much more complete and comprehensively indexed collection; see Charlotte Garside, *Women in Chancery: An analysis of Chancery as a court of redress for women in seventeenth-century England* (Unpublished PhD Thesis: University of Hull, 2019), pp. 13, 24. Garside notes that in a medieval context, Timothy Haskett has also complained of ‘too much rather than too little’ Chancery material; ‘The medieval English Court of Chancery’, *Law and History Review*, 14 (1996), pp. 245–313.

³⁵ Reid and van de Wetering, ‘Conservation’, pp. 189-90. For more on the ‘salvage operation’ which followed the 1922 fire, see Peter Crooks and Zoë Reid with Jessica Baldwin, Ciara Fahy, and Brian Gurrin, ‘The salved records of 1922’, *Analecta Hibernica*, 53 (2023), pp. 17–26.

³⁶ The only exception being NAI, CP/BB/205, the Bill of Sebilla Dillon on behalf of herself and the rest of the children of Henry Dillon late of Kentstown, Co. Meath v. Bartholomew Kent, 23 November 1622. Bartholomew

Kent's Answer can be found at NAI, BB/216, dated 27 November 1622, which says something of the speed of proceedings in Irish Chancery.

³⁷ See NAI, CHAN/Ancient Pleadings, Index 1617-1624; CHAN/Ancient Pleadings, Index 1624-1629; CHAN/Ancient Pleadings, Index 1629-1634; CHAN/Index Ancient Answers 1569-1618. We are grateful to Dr Brian Donovan for alerting us to the significance of the Chancery Indexes, which he outlines in 'Why the Chancery Pleadings survived 1922', *Archivium Hibernicum* (forthcoming, 2026).

³⁸ O'Dowd, 'Women and the Irish Chancery', p. 472.

³⁹ O'Dowd, 'Women and the Irish Chancery', p. 471.

⁴⁰ Some traces of this material can be found amongst the collection, including interrogatories, inventories and accounts (forms of exhibit), and perhaps some depositions in unlisted boxes which have yet to be conserved.

⁴¹ Some examples include, Maria Cioni, *Women and law in Elizabethan England, with particular reference to the Court of Chancery* (Garland: New York, 1985); Amanda Capern, 'Maternity and Justice in the Early Modern English Court of Chancery', *Journal of British Studies*, 58 (2019), pp. 701-16 and 'Emotions, Gender Expectations, and the Social Role of Chancery, 1550-1650' in Susan Broomhill (ed.), *Authority, Gender and Emotions in Late Medieval and Early Modern England* (Palgrave Macmillan: London, 2015), pp. 187-209; Amy Erickson, *Women and property in early modern England* (Routledge: London, 1995); Elise Dermineur (ed.), 'Women and Money: Credit, Debt, and Status in the Eighteenth-century London Court of Exchequer' in *Women and Credit in Pre-industrial Europe* (Brepols: Turnhout, 2018), pp. 281-99; Daniel Patterson, 'The Story of Lucy Browne: Women's Agency, "Voices" and the Evidence of Chancery Depositions', *Women's History Today*, 3 (2021), pp. 30-39.

⁴² The pleadings themselves, particularly the bills of complaint, rarely include the official titles of the suits, and sometimes the names of the primary litigants are unclear or missing entirely. For reasons which are unclear, answers more frequently include the titles of lawsuits.

⁴³ See, for instance, NAI CP/I/62 and CP/J/143, both cited by Mary O'Dowd as cases where fathers sued former sons in law for the return of marriage goods in the event of separations, leaving the women concerned out of the lawsuit; 'Women and the Irish Chancery Court', p. 474.

⁴⁴ For more on the pioneering work of Emily Kadens, see <<https://www.transkribus.org/success-story/reading-english-secretary-hand-with-ai>> accessed 21 October 2024.

⁴⁵ NAI, CP/B/ 231, Bill of Conner McDonagh of Ballindoon, Co. Sligo, gent. and Una ny Rourke alias Dogherty alias ny Donagh his wife v. Sir Cahaire O Dogherty, knt., undated; NAI, B/60, Bill of Ann Cowenton of Dublin v. John Sheirvan of the Grandge, Co. Dublin, farmer, undated; NAI, BB/205, Bill of Sebilla Dillon on behalf of herself and the rest of the children of Henry Dillon late of Kentstown, Co. Meath v. Bartholomew Kent, 23 November 1622.

⁴⁶ NAI, CP/N/35; H/160; I/246; B/112; T/84.

⁴⁷ Mary O'Dowd characterises the fragmentary nature of surviving sources for Irish social history as 'the dispersed archive', Mary O'Dowd, 'Men, women, children and the family, 1550-1730' in Jane Ohlmeyer (ed.), *The Cambridge history of Ireland: II, 1550-1730* (Cambridge, 2018), pp. 299-300.

⁴⁸ A collection of will books, containing copies of original wills proved in the Prerogative Court of Armagh during the period c.1664-1729 held in the NAI is currently being processed by the project. The originals perished in the 1922 destruction of the Irish Public Records Office.

⁴⁹ For a discussion of wills as mediated sources, see Tait, 'Writing the social and cultural history of Ireland, 1550-1660', pp. 27-48.

⁵⁰ For detailed discussion see Harry Smith and Emily Vine, 'Material and digital archives: the case of wills', *Transactions of the Royal Historical Society* (advance access, 2024), pp. 16-17. [doi:10.1017/S0080440124000094]

⁵¹ 'The Material Culture of Wills' project is harnessing Transkribus to transcribe 25,000 wills held in the PCC, <https://sites.exeter.ac.uk/materialcultureofwills>, accessed 18 October 2024.

⁵² The volumes date from 1384 to 1858 and contain hundreds of thousands of registered copies of wills.

⁵³ https://virtualtreasury.ie/search-results?totalElementsInt=100&pageNumberInt=0&kwList=TNA%20PROB&kwOperList=ALL&enableQueryExpansion=false&enableFuzzySearch=false&searchContentDate_begin=1200-01-01&searchContentDate_end=1870-12-31&kwSearchFieldList=referenceCode&resultSorting=relevance, accessed 26 October 2024.

⁵⁴ Will of Magdalen Ring or Magdalene Ringe, Singlewoman of Marolopp, Waterford, TNA, PROB 11/144/117.

⁵⁵ Jonathan Barry and Christopher Brooks (eds), *The middling sort of people: culture, society and politics in England 1550-1800* (Basingstoke, 1994). For categorisation of social status in early modern England: <https://middlingculture.com/social-statuses-of-early-modern-england/>

-
- ⁵⁶ For a recent study of the engagement of married women in testamentary practice in late medieval England see Cordelia Beattie, 'Married women's wills: probate, property, and piety in later medieval England', in *Law and History Review* 37(1) (2019), pp. 29-60. [doi:10.1017/S0738248018000652]
- ⁵⁷ Will of Susannah Feilding alias Booth of Dublin, County Dublin, TNA, PROB 11/483/262.
- ⁵⁸ Prell 'Early modern letters as a challenge automated handwriting recognition'.
- ⁵⁹ 'Introducing Transkribus Super Models – Get Access to 'The Text Titan I' <https://readcoop.eu/introducing-transkribus-super-models-get-access-to-the-text-titan-i>, accessed 26 October 2024.
- ⁶⁰ Harry Smith, 'Digitising 25,000 wills: method and accuracy', Material Culture of Wills Project Blog, <https://sites.exeter.ac.uk/materialcultureofwills/2024/04/09/digitising-25000-wills-method-and-accuracy/>, accessed 26 October 2024.
- ⁶¹ Smith and Vine, 'Material and digital archives: the case of wills', p. 20.
- ⁶² Smith and Vine, 'Material and digital archives: the case of wills', p. 27.
- ⁶³ Will of Isabella Walker of County Tyrone, TNA, PROB 11/489/62.
- ⁶⁴ Will of Mary De St. Julien Adlerson or Adlerson, TNA, PROB 11/493/432; will of Judith Jaquean, TNA, PROB 11/454/355.
- ⁶⁵ Clodagh Tait has discussed the 'cultural ambidexterity' of testators in early modern Ireland, 'Writing the social and cultural history of Ireland', p. 34.
- ⁶⁶ Will of Frances Parker of County Carlow, TNA, PROB 11/162/207.
- ⁶⁷ There are currently two Irish-language models available on Transkribus, trained chiefly on printed text. For a discussion see Deirdre Ní Chonghaile, Oksana Dereza and Nicholas Wolf, '#AnGadhal: Training the first Irish-English bilingual OCR model', Aug. 2024, <https://hardimanlibrary.blogspot.com/2024/08/an-gadhal-ocr.html>, accessed 26 October 2024.
- ⁶⁸ Qiuzi Guo, 'Prompting change: ChatGPT's impact on digital humanities pedagogy – a case study in art history', *International Journal of Humanities and Arts Computing*, 18:1 (2024), pp. 58–78 [https://doi.org/10.3366/ijhac.2024.0321]
- ⁶⁹ 'Welcome to the 'Walled Garden.' Is this education's solution to AI's pitfalls?', *Education Week*, July 2023, <https://www.edweek.org/technology/welcome-to-the-walled-garden-is-this-educations-solution-to-ais-pitfalls/2023/07>, accessed 26 October 2024.