



Digitising Death: Benchmarking Genealogical Data and Recovering Women's Histories in Early Modern Ireland

DISCUSSION PAPER

BRONAGH ANN MCSHANE

DIEGO RINCON-YANEZ

FELIX VANDEN BORRE

JANE OHLMEYER

DECLAN O'SULLIVAN

*Author affiliations can be found in the back matter of this article

ju[ubiquity press

ABSTRACT

This article examines how benchmarking digital methods can advance historical research by recovering women's lives from fragmented and underused archives. It focuses on the Funeral Entries held in the Genealogical Office of the National Library of Ireland, a rich but understudied manuscript collection compiled by the Ulster King of Arms between the late sixteenth and early eighteenth centuries. These records, which document death dates, kinship networks, and social affiliations, contain an unusually high proportion of women for the early modern period (about 38 percent), offering rare insight into gendered experiences of death, family, and memory.

The article emerges from the ERC-funded VOICES project, which develops AI-powered approaches to recover women's voices and experiences from early modern sources (VOICES Project, 2023; <https://voicesproject.ie>). As part of this work, we present a benchmarking experiment using the Funeral Entries to assess the ability of Handwritten Text Recognition (HTR) and Named Entity Recognition (NER) models to cope with early modern orthography, multilingual naming practices, and manuscript variability. These experiments not only illuminate the constraints of existing tools but also demonstrate how benchmarking can generate reusable workflows, inform the creation of annotated gold standards, and support the production of FAIR-aligned humanities data.

Rather than offering a completed dataset, the article argues for the historiographical value of iterative benchmarking. We show how evaluating and refining computational methods reframes questions of archival visibility, evidentiary status, and the interpretive potential of genealogical records, positioning the Funeral Entries not merely as instruments of male lineage but as essential sources for recovering early modern women's social worlds.

CORRESPONDING AUTHOR:

Bronagh Ann McShane

School of Histories and
Humanities, Trinity College
Dublin, Ireland

mcshaneb@tcd.ie

KEYWORDS:

digital humanities;
benchmarking; named entity
recognition; early modern
Ireland; women's history; FAIR
data

TO CITE THIS ARTICLE:

McShane, B. A., Rincon-Yanez, D., Vanden Borre, F., Ohlmeyer, J., & O'Sullivan, D. (2026). Digitising Death: Benchmarking Genealogical Data and Recovering Women's Histories in Early Modern Ireland. *Journal of Open Humanities Data*, 12: 45, pp. 1–13. DOI: <https://doi.org/10.5334/johd.491>

(1) CONTEXT AND MOTIVATION

VOICES (Voices of Women in Early Modern Ireland, 1550–1700) is a five-year ERC-funded project that brings together historians, digital humanists, and computer scientists to ask how digital methods might recover the lived experiences of non-elite women in a period marked by war, religious conflict, colonisation, and displacement (VOICES Project, 2023-; Ohlmeyer, 2018, 2023). The questions it addresses are as much methodological as they are historical. Placing AI-powered tools into a sustained conversation with archives that were never designed to record women’s voices compels us to examine both how data is constituted and how analytical frameworks may reinscribe or counteract older patterns of exclusion (Mac Curtain and O’Dowd, 1991; O’Dowd, 2005; Nolan and McShane, 2022).

Early modern archives offer an especially revealing testing ground for these questions. Even in rich collections, women are more often inferred than explicitly named, glimpsed as wives, mothers, daughters, widows, or “relicts” rather than as unambiguously autonomous subjects (Palmer, 2018; Tait, 2022). The VOICES project engages with several major archival corpora that exemplify these challenges while also demonstrating the possibilities of digital recovery. The 1641 Depositions, eyewitness testimonies gathered in the aftermath of the Irish rebellion of 1641, include women’s statements at a rate of roughly one in eight (1641 Depositions Project, 2010, <https://1641.tcd.ie>). The Prerogative Court of Armagh wills, which governed testamentary business for individuals with property in more than one diocese, include women as testators in roughly one in nine cases (McShane, 2025). Meanwhile, the Chancery pleadings, produced by the equity court that mediated disputes over land, debt, and inheritance, show a slightly higher level of female participation, with about one in five cases involving women as plaintiffs (O’Dowd, 1999). Taken together, these sources illustrate both the extent and the limits of women’s archival visibility. They capture women at moments of legal, social, or personal rupture, yet they do so in ways shaped by the gendered assumptions of contemporary bureaucracies. Reconstructing women’s lives, and the social worlds through which they moved, therefore requires reading across documents, genres, and repositories. It also demands the ability to scale such reading beyond the limits of manual transcription and unaided search (Mac Curtain et al., 1992).

Recent advances in handwritten text recognition (HTR) and natural language processing (NLP) (Navigli, 2018) offer precisely that potential. Yet the same tools that seem to promise scale, often bring with them an implicit normativity. Most are trained on modern corpora, stabilised by standardised orthography, and optimised for contemporary genres of writing that bear little resemblance to the abbreviatory habits and relational semantics of early modern documentary culture (Piotrowski, 2012; Sadek et al., 2024). The resulting domain shift raises questions not only about accuracy but also about who becomes visible in the data. Benchmarking, the systematic evaluation of computational models against declared standards, provides a means to confront these issues directly. It enables researchers to quantify errors, for instance, to identify where women’s names or Irish-language forms are lost or distorted, and to iteratively improve model performance while maintaining interpretive awareness (Tual et al., 2023; Provatorova et al., 2020).

This article presents an extended case study from VOICES: our work on the Funeral Entries preserved in the Genealogical Office of the National Library of Ireland (National Library of Ireland, n.d.; McAnlis, 2014; McShane and Vanden Borre, 2025). The discussion traces the transformation from manuscript facsimile to processable text, from transcription to entity recognition, and, later, from entities to a knowledge graph (Rincon-Yanez & Senatore, 2022) capable of modelling kinship and place across collections (Randles et al., 2024; Yaman et al., 2024) (see, for example, the Virtual Record Treasury of Ireland initiative; <https://kg.virtualtreasury.ie>). At each stage, we show how benchmarking shapes both technical choices and historical interpretation. Rather than treating benchmarking as an external quality-control procedure, we argue that it functions as a core historical method, a practice of iterative scrutiny that exposes the assumptions embedded in both archival sources and computational models (Nolan and McShane, 2022; McShane and Vanden Borre, 2025).

Because the Funeral Entries record women with a frequency and granularity that is rare in early modern Irish sources (MacCurtain and O’Dowd, 1991; O’Dowd, 2005; Tait, 2022), they provide a crucial case study for assessing whether digital methods can do justice to the gendered dimensions of the past.

(1.1) BEYOND BUREAUCRATIC RECORDS: THE FUNERAL ENTRIES

The Funeral Entries span seventeen volumes and more than three thousand manuscript pages and extend from the late sixteenth to the early eighteenth century ([National Library of Ireland, n.d.](#); [McAnlis, 2014](#)). Compiled by the Ulster King of Arms, they were intended as bureaucratic instruments: tools for verifying genealogical claims and maintaining heraldic order within the colonial elite ([Ohlmeyer, 2018](#)). Yet because lineage and inheritance lay at their core, women appear throughout the collection as daughters, wives, mothers, and widows. Their presence is relational, but that relationality is not trivial; it is central to how these records construct social continuity and legitimacy. A gendered reading of the Funeral Entries highlights the active role women played in linking families, offices, and regions through patterns of marriage, descent, and kinship ([Mac Curtain and O'Dowd, 1991](#); [O'Dowd, 2005](#)).

The entries are succinct yet densely layered. Many include a date and place of death, a burial site, and a concise account of family connections, husbands and former husbands, children, in-laws, and significant kin. At times, they evoke the cosmopolitan texture of early seventeenth-century Dublin, where marriages linked Old English officials to continental artisans. At other moments, they expose the social negotiations of confessional division, as unions bridged Catholic and Protestant households ([Ohlmeyer, 2018, 2023](#); [McShane, 2019](#)). Occasionally, the bureaucratic façade softens and emotion surfaces: a woman is buried with most of her issue, or her three children have died in infancy. These brief asides remind us that these documents are not only instruments of lineage but also records of grief and continuity, capturing the affective dimensions of family life that underpinned genealogical logic ([O'Dowd, 2005](#)).

For historians of women, the scale of inclusion is remarkable. A preliminary survey of one volume (GO MS 66) shows that roughly 38 percent of entries concern women between 1600 and 1620, with the broader series likely approaching two in five. Such proportions are extraordinary by early modern standards, where women's appearances in state or legal records are often sporadic and contingent upon litigation, deviance, or crisis ([Mac Curtain and O'Dowd, 1991](#); [Palmer, 2018](#); [Tait, 2022](#)). The Funeral Entries provide a more systematic, if socially selective, view of women's lives, offering data that support both individual reconstruction and quantitative analysis ([Mac Curtain et al., 1992](#); [Nolan and McShane, 2022](#); [McShane and Vanden Borre, 2025](#)).

This combination of scale and consistency distinguishes the Funeral Entries from other core sources used in Irish women's history. The 1641 Depositions, wills, and Chancery pleadings tend to capture women in moments of rupture, when property is disputed, when violence disrupts communities, or when death necessitates formal arrangements of goods ([O'Dowd, 2005](#); [Nolan, 2020](#)). The Funeral Entries, by contrast, record death as a routine genealogical event. They allow us to see women not only in exceptional circumstances but as integral to the ordinary reproduction of family and status. The careful noting of successive marriages, who married whom, in what order, and with what issue, enables historians to trace how women's marital alliances helped sustain civic, confessional, and regional elites over time ([MacCurtain and O'Dowd, 1991](#); [Ohlmeyer, 2018](#)).

The source is also revealing in its treatment of ethnicity and confession. Early modern Ireland encompassed Gaelic Irish, Old English, New English, and smaller migrant groups such as Huguenots and Italian artisans ([Ohlmeyer, 2018, 2023](#)). The Funeral Entries reflect this diversity, recording women as daughters of Gaelic and Old English families, wives of New English officials, and sometimes as figures whose names point to continental origins. These entries offer a microcosm of cultural negotiation, showing how confessional and ethnic identities were redefined through marriage and memorialisation, and capturing the paradox of Catholic women memorialised within a heraldic record administered by a Protestant state ([Mac Curtain and O'Dowd, 1991](#); [O'Dowd, 2005](#); [Palmer, 2018](#)).

Because the entries note burial places in addition to kinship ties, they open further insight into women's spatial and religious worlds. Patterns of burial in specific churches or vaults may reflect denominational preference, local attachment, or continued association with family estates ([O'Dowd, 2005](#)). When read across generations, these records trace persistence and change, familial clusters, shifts in confessional allegiance, or patterns of mobility as women marry across regions. For historians interested in women's engagement with spatial and devotional life, the Funeral Entries offer unusually concrete coordinates ([Tait, 2022](#)).

The language of reproduction that recurs throughout the manuscripts adds yet another dimension. Phrases such as “without issue”, “leaving no issue”, or “having many children” illuminate how fertility and lineage structured early modern understandings of value and continuity. When linked to age, marital sequence, or notes on deceased children, these formulae permit cautious reconstruction of reproductive histories, rarely possible elsewhere in Irish material (O’Dowd, 2005; Palmer, 2018). Even in their brevity, they expose how women’s bodies and reproductive capacities were woven into the genealogical imagination.

The Funeral Entries matter, then, not simply because they record women more often than other sources, but because they situate women at the intersection of family strategy, colonial governance, religious identity, and emotional life (Mac Curtain and O’Dowd, 1991; Ohlmeyer, 2018, 2023; Nolan and McShane, 2022). Precisely because of this interpretive richness, the methods used to digitise and analyse the Funeral Entries directly shape the histories that can be written from them. If women’s names are mistranscribed or misrecognised, or if Irish-language patronymics are ignored by NER models, the very women made visible by the manuscript risk slipping back into digital invisibility (Sadek et al., 2024; Tual et al., 2023; Provatorova et al., 2020; McShane and Vanden Borre, 2025). By placing benchmarking at the centre of our workflow, we aim not only to unlock the evidential potential of the Funeral Entries but also to ensure that digital tools expand rather than constrain the field of vision (Wilkinson et al., 2016; Nolan and McShane, 2022).

(2) DATASET DESCRIPTION

REPOSITORY LOCATION

<https://doi.org/10.5281/zenodo.18174069>

REPOSITORY NAME

Zenodo

OBJECT NAME

VOICES-FuneralEntries-NER-0.0.3

FORMAT NAMES AND VERSIONS

The repository (version 0.0.3) contains multiple transcription iterations of the first three folios of NLI GO MS 66 (.csv, .ods), Named Entity Recognition results (.json), and validated results (.csv) along with documentation and data curation guides (.md) and Notebooks used in the creation of the dataset (.ipynb).

CREATION DATES

From 2025-08-11 to 2025-10-07.

DATASET CREATORS

The dataset creators are Bronagh Ann McShane (transcriptions), Diego Rincon-Yanez (NER curation), and Felix Vanden Borre (transcriptions, NER validation).

LANGUAGES

English (early modern).

LICENSE

Creative Commons Attribution 4.0 International (CC BY 4.0)

PUBLICATION DATE

2026-01-07.

(3) METHOD

(3.1) FROM MANUSCRIPT SOURCES TO PROCESSABLE TEXT

At the outset of this work, the Funeral Entries manuscripts existed as digital images on the National Library of Ireland's website but were not text searchable. A first challenge in using computational methods on this historical source has therefore been to transform digital facsimiles into processable text, using the HTR software Transkribus (READ-COOP, 2025). Transkribus was designed for the transcription of historical documents and helps to expedite what would otherwise be a labour-intensive component of the research process. Users can mark up the specific regions of a page from which transcriptions should be generated.

The material intricacies of the Funeral Entries complicated this first step. A single page can contain anywhere between one and thirteen entries. Many entries are accompanied by additional information: a heading that contains the surname of the deceased, marginal notes, a heraldic shield, or some combination of these elements.

To manage this complexity, we designed a tagging protocol to keep each entry separate while correctly capturing all the information required. Transkribus enables users to label text regions with a set of tags that can be activated or customised at any time. We distinguished individual Funeral Entries by creating at least one heading region and one paragraph region per entry, with the paragraph region containing the main body of text (Figures 1 & 2). Where one or more marginalia existed, these were also captured and labelled accordingly. The tagged regions could then be parsed into individual entries after export from Transkribus in PAGE-XML format. Because drawings or symbols such as coats of arms can interfere with text recognition, we avoided capturing them when marking up text regions.

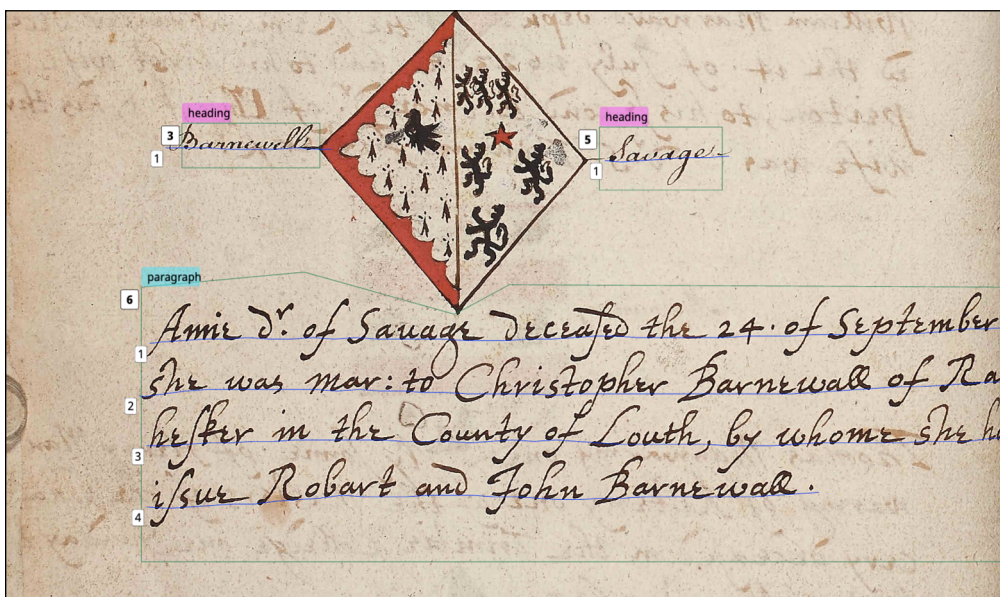


Figure 1 Transkribus tagging of GO MS 68 fol. 30v (funeral entry of Amy Savage) showing double headers and paragraph region.

One fundamental material characteristic of the manuscripts that this system cannot fully address is the presence of several different hands within the corpus. The early modern period saw the development of multiple handwriting styles and typographical practices, which varied across countries, contexts, and individual scribes. Several hands sometimes coexist within a single volume, and the poor quality of some entries suggests that, amid the turmoil of seventeenth-century Ireland, speed sometimes took precedence over clarity.

Transkribus contains a range of open-access models that can be used by researchers and also allows users to train their own model based on a specific source or group of sources. Because of the brevity of the Funeral Entries and the constant variation in handwriting, we did not have enough homogeneous training data to build a stable custom model, nor was this feasible within the time constraints of the project. Instead, we opted to use one of the platform's most powerful super models, Text Titan I (READ-COOP, 2023).

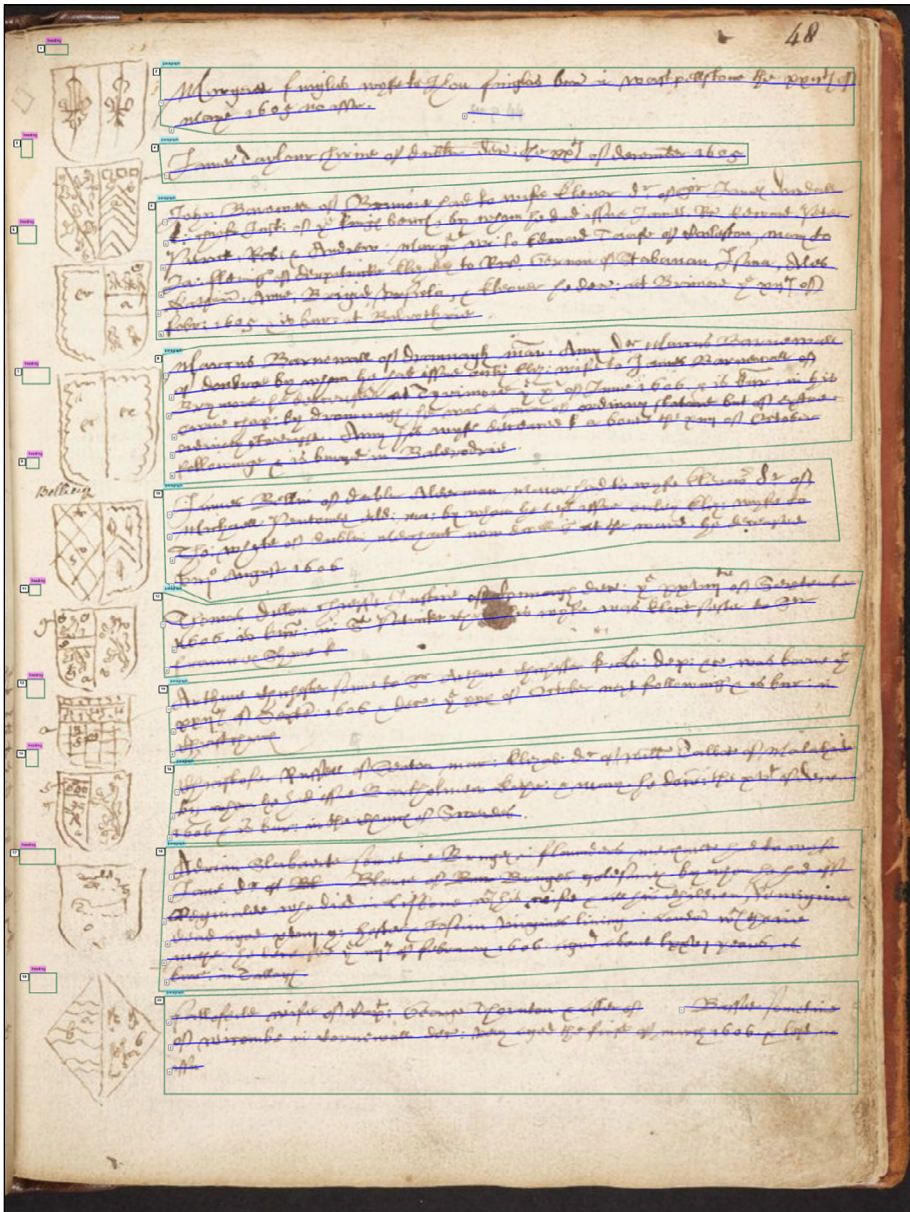


Figure 2 Transkribus tagging of GO MS 64 fol. 48r showing empty headers being used to separate individual entries.

Even this state-of-the-art HTR model has struggled with aspects of the manuscripts. The presence of Irish names and place names throughout the corpus means that the transcription process would benefit from a model trained on historical Irish-language data. Together, these factors have resulted in some excellent, and also some disappointing transcriptions.

The funeral entry of Mary Lady Tyrell illustrates a relatively accurate output:

“Mary Lady Tyrell, dr. of Walter segrave alderman maior of Dublin, & wife to Sir John Tyrell Knight ald: and maior of the same cittie: deceased in December 1604. and is, buried in St. Audoens, church. Shee had issue walter, Mathew Richarde, Edward, Cicilie (wife to Nichas Stevens, some time shireve of Dublin) Marg. and three more deade in there infaucie.” (NLI GO MS 66 fol. 1v)

Aside from the Segrave family name not being capitalised, the main HTR errors are the misspelling of infaucie and an intrusive comma after St. Audoens. The overall transcription, however, preserves the historical spelling and typography sufficiently to allow identification of key entities and relations.

A second example, the entry for Alice Segrave, demonstrates the limits of the model:

“Also dt. of waster Segraue of Dublin Alderman Maior & wi: of John eldest sonne of Richard Pagan Ald: n2. of sy samie, deceased the xym. of may grs. DE. viiq. & is but: in St. Andoeno, church. Shee was mother to Richard, Thomas, George & John: cicilie, Mary & Anne.” (NLI GO MS 66 fol. 11v)

The first sentence should read: “Ales dr. of Walter Segrave of Dublin Alderman Maior & wi: of John eldest sonne of Richard Fagan Ald: M. of the same, deceased the xxiith of May MDCVIII and is bur: in St. Audoens Church.” It is easy to miss that Also is a misrendering of Ales (Alice) or that Pagan should read Fagan. Roman numerals are also particularly challenging for HTR.

Our transcriptions are therefore processable but uneven in quality. For the purposes of this article, the crucial question is how such variation, encompassing historical spellings, Irish-language forms, early modern typography, and HTR errors, affects the performance of off-the-shelf NER models, and how benchmarking can make these effects visible and addressable.

(3.2) GENERATING MULTIPLE TRANSCRIPTION VARIANTS FOR BENCHMARKING

To investigate how transcription quality affects NER performance, we created multiple versions of a small, carefully controlled subset of the Funeral Entries. We selected the first twelve entries in GO MS 66 and generated a series of transcriptions that held content constant while varying levels of normalisation, correction, and interpretive intervention.

Five transcription variants were produced:

- **Manual transcription (MAN):** A historian-generated version retaining early modern spellings while correcting obvious errors.
- **EyeCR transcription (EYE):** Produced using an internal transcription tool that can connect to different LLMs, developed at the ADAPT Centre. For our case study, the LLM that was connected to EyeCR was GPT-4.1-mini model.
- **Transkribus transcription (TRA):** Raw HTR output exported directly from Transkribus.
- **Curated Transkribus transcription (CUR):** A diplomatically curated version of TRA, correcting clear HTR errors while preserving original orthography.
- **Transkribus + VARD transcription (TKV):** The TRA version processed with VARD2 to normalise historical spelling variants. VARD2 is an interactive tool designed to deal with spelling variation in historical corpora (VARD, 2025).

An additional manual transcription was created and annotated with named entities. This manually annotated set acted as an approximate ground truth against the transcriptions used for benchmarking, recognising that no transcription can be fully authoritative given the interpretive decisions required.

Curating this benchmarking dataset, including the ground truth, highlighted how historical and manuscript realities resist straightforward computational encoding. The funeral entry of James Barnewall, for example, mentions the parish of Stabannan in County Louth. In the manuscript, the scribe splits the place name across two lines and inserts a colon to indicate continuation, writing it as “Staba: nan” (Figure 3). The same device appears in “Ro: bart”. In both our manual transcriptions and the NER ground truth, these were joined into “Stabanan” and “Robart” (Robert). This decision makes sense from the standpoint of data regularisation but means that the ground truth is not, in a strict palaeographic sense, identical to the manuscript. Creating a benchmarking dataset thus made visible the interpretive steps required to turn manuscripts into machine-readable data and forced us to reflect on how such steps affect downstream model performance.

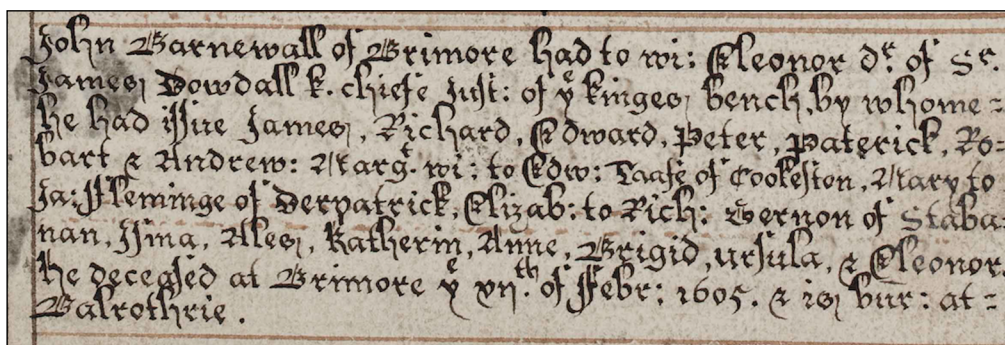


Figure 3 Funeral Entry of James Barnewall (NLI GO MS 66 fol. 3v).

(3.3) NER MODELS AND CONFIGURATIONS

Within VOICES, reproducible and scalable NLP workflows are essential, both because of the heterogeneity of the manuscript collection and because the project ultimately aims to construct a knowledge graph capable of linking persons and places across archival corpora (Debruyne et al., 2022). To support this, we implemented an experimentation pipeline to evaluate off-the-shelf NER models on the transcription variants described above.

In the benchmark experiment, each paragraph of a Funeral Entry was processed using two standard NER approaches: a transformer-based BERT model (Devlin et al., 2019) and the CRF-based Stanford CoreNLP (Qi et al., 2020) tagger (Figure 4). Their outputs were then manually annotated to quantify model performance using the synthetic metric outlined in Section 2.4.

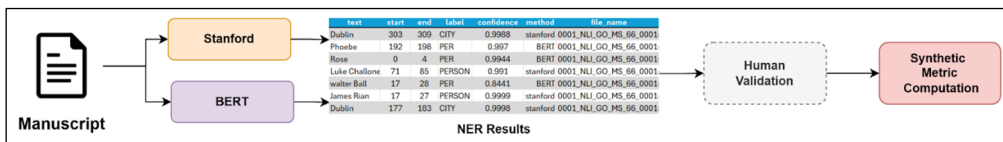


Figure 4 High-Level experiment setting with experiment setting.

Because one of our key aims was to assess how transcription quality shapes NER results, we ran each model across all five curated transcription variants (MAN, EYE, TRA, CUR, and TKV). We also tested a combined configuration that merges predictions from BERT and Stanford, retaining only those with a confidence score of 70 percent or higher and resolving overlapping spans consistently. The full structure of the experiment, including inputs, outputs, and expected behaviour, is shown in Figure 5.

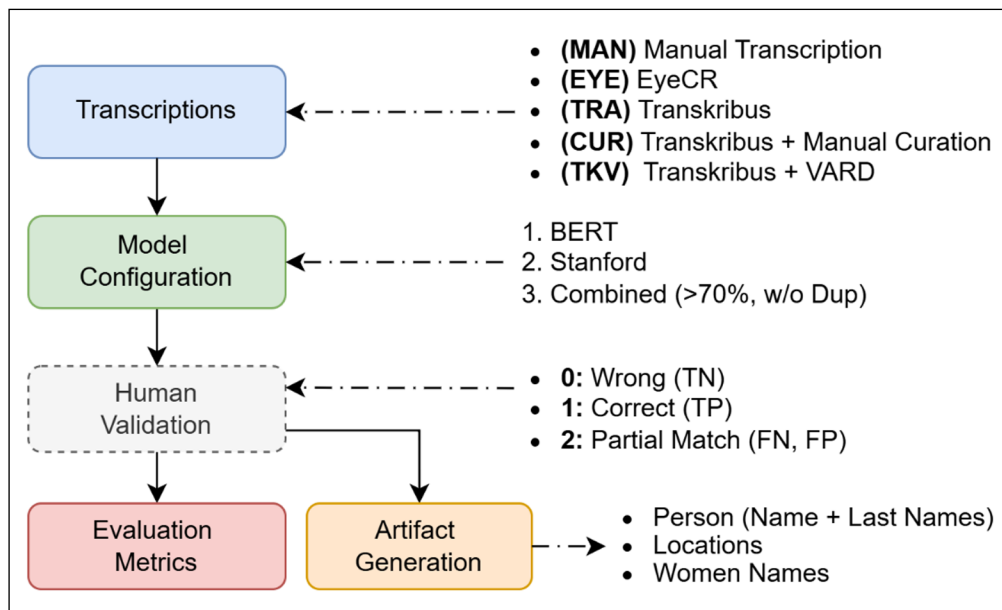


Figure 5 Experiment detail with inputs and expected experiment outputs.

The experiment evaluates three NER configurations: a BERT-based model (see Table 1), the Stanford CoreNLP tagger (see Table 2), and a combined scenario in which we take the union of predictions from both systems but retain only those with a confidence of 70 percent or higher (see Table 3). In this combined scenario, overlapping spans are resolved so that entities detected by BERT do not contradict those produced by Stanford. A detailed version of the experiment scenarios, inputs and expected results can be seen in Figure 5.

(3.4) EVALUATION PROTOCOL AND SYNTHETIC METRICS

For each entry, and for each model and transcription combination, the output entities were manually annotated using a simple three-level scheme: 0 for incorrect, 1 for correct, and 2 for partial match. This annotation process produced a small but carefully controlled benchmark suited to our particular corpus and research questions.

FILE	TRA		EYE		CUR		TKV		MAN	
	E	SA	E	SA	E	SA	E	SA	E	SA
0001r	4	100.00%	4	100.00%	3	33.33%	4	100.00%	6	83.33%
0001r	5	80.00%	7	57.14%	6	83.33%	5	80.00%	7	85.71%
0001v	10	70.00%	8	87.50%	13	69.23%	11	72.73%	13	84.62%
0001v	6	50.00%	5	80.00%	6	100.00%	6	83.33%	5	80.00%
0002r	3	100.00%	3	100.00%	3	66.67%	3	100.00%	3	100.00%
0002r	11	63.64%	22	77.27%	13	84.62%	12	83.33%	11	72.73%
0002v	14	28.57%			14	42.86%	13	38.46%	13	84.62%
0002v	5	20.00%	4	25.00%	4	75.00%	5	20.00%	4	75.00%
0003r	6	66.67%	6	83.33%	7	100.00%	6	100.00%	6	100.00%
0003r	10	40.00%	11	36.36%	9	44.44%	9	55.56%	10	60.00%
0003v	1	100.00%	2	50.00%	2	50.00%	2	50.00%	2	50.00%
0003v	19	68.42%	22	72.73%	28	92.86%	24	79.17%	23	78.26%
Avg		65.61%		69.94%		70.19%		71.88%		79.52%

Table 1 Stanford Results organised by overall best (right), overall worst(left).

FILE	TRA		EYE		TKV		MAN		CUR	
	E	SA	E	SA	E	SA	E	SA	E	SA
0001r	8	75.00%	8	75.00%	8	75.00%	8	75.00%	7	71.43%
0001r	7	57.14%	6	66.67%	7	57.14%	6	100.00%	6	83.33%
0001v	11	54.55%	11	72.73%	11	81.82%	12	91.67%	13	92.31%
0001v	11	36.36%	11	45.45%	9	66.67%	9	77.78%	8	100.00%
0002r	3	100.00%	3	100.00%	3	100.00%	3	100.00%	3	100.00%
0002r	13	53.85%	31	61.29%	12	58.33%	12	50.00%	13	84.62%
0002v	21	23.81%			16	43.75%	21	61.90%	18	44.44%
0002v	11	36.36%	6	33.33%	10	30.00%	6	66.67%	6	66.67%
0003r	7	42.86%	8	50.00%	6	83.33%	7	85.71%	7	100.00%
0003r	10	70.00%	11	54.55%	10	50.00%	14	42.86%	9	77.78%
0003v	1	100.00%	2	50.00%	2	100.00%	2	100.00%	2	100.00%
0003v	30	56.67%	30	63.33%	31	67.74%	33	66.67%	29	96.55%
Avg		58.88%		61.12%		67.82%		76.52%		84.76%

Table 2 BERT Results organised by overall best (right), overall worst(left).

FILE	TRA		EYE		TKV		CUR		MAN	
	E	SA	E	SA	E	SA	E	SA	E	SA
0001r	5	100.00%	5	100.00%	6	83.33%	6	83.33%	5	100.00%
0001r	7	57.14%	6	50.00%	5	60.00%	5	80.00%	6	83.33%
0001v	13	53.85%	9	66.67%	11	72.73%	14	71.43%	12	83.33%
0001v	11	54.55%	5	80.00%	9	66.67%	8	87.50%	7	85.71%
0002r	3	100.00%	3	100.00%	3	100.00%	3	66.67%	3	100.00%
0002r	17	70.59%	30	60.00%	17	64.71%	13	76.92%	16	50.00%
0002v	27	37.04%			17	52.94%	18	50.00%	22	63.64%
0002v	22	50.00%	5	40.00%	10	20.00%	5	80.00%	6	66.67%
0003r	14	57.14%	8	62.50%	7	85.71%	6	100.00%	7	85.71%
0003r	35	65.71%	15	40.00%	11	54.55%	10	70.00%	11	63.64%
0003v	10	50.00%	2	50.00%	3	66.67%	3	66.67%	3	66.67%
0003v	9	44.44%	25	68.00%	28	75.00%	27	96.30%	17	94.12%
Avg		61.71%		65.20%		66.86%		77.40%		78.57%

Table 3 Combined results organised by overall best (right), overall worst(left).

Standard NER evaluation relies on counts of true positives, false positives, true negatives, and false negatives, together with derived measures such as precision, recall, and F1 score. In newly digitised humanities collections, however, a robust gold standard rarely exists, and the number and type of candidate entities varies across transcriptions. For this reason, we adopted a synthetic evaluation protocol.

Let N_1 denote the number of entities annotated as correct, N_2 the number annotated as partial matches, and N_0 the number annotated as incorrect. The total number of evaluated entities is then:

$$N_{Total} = N_1 + N_2 + N_0$$

Under this annotation pattern, a strict and relaxed accuracy can be computed as the number of correct entities divided by the total number of evaluated entities, multiplied by 100 to give a percentage:

$$Strict\ Accuracy\ (SA)\ (\%) = \frac{N_1}{N_{Total}} = \frac{N_1}{N_1 + N_2 + N_0} * 100$$

$$Relaxed\ Accuracy\ (RA)\ (\%) = \frac{N_1 + N_2}{N_{Total}} = \frac{N_1 + N_2}{N_1 + N_2 + N_0} * 100$$

This simple metrics does not capture all nuances of entity recognition, but it provides a transparent and reproducible way to compare model performance across different transcriptions and configurations. Importantly, the annotation-based approach allows us to work with multiple plausible transcriptions rather than assuming the existence of a single, fully authoritative ground truth.

(4) RESULTS AND DISCUSSION

Across the three NER scenarios and the five transcription versions, several patterns emerge. Overall, the Transkribus-only (TRA) and EyeCR (EYE) versions achieve the lowest strict accuracy scores. By contrast, the Transkribus plus VARD (TKV) and curated Transkribus (CUR) versions obtain the highest scores, with a gap of almost fourteen percentage points between the poorest and the best overall results.

In the BERT-only scenario (see [Table 2](#)), the versions involving manual intervention (MAN, CUR, and TKV) consistently outperform those relying solely on automatic transcription. The improvement relative to the Stanford-only scenario is particularly marked, with BERT achieving roughly twice the strict accuracy in some configurations. In the combined scenario, which unions predictions from BERT and Stanford with a 70 percent confidence threshold and resolves duplicates, the results closely track those of BERT alone, ranking the collections in the same order (TKV, CUR, MAN, EYE, TRA).

Two findings are especially significant from a benchmarking perspective. First, historical manuscript material still requires human-in-the-loop processing to achieve competitive performance. Fully unattended pipelines underperform relative to hybrid workflows that incorporate expert curation at one or more stages. Second, the difference in performance between the fully manual transcription and the curated Transkribus version is relatively small (around 1.1 percent in our scenario), suggesting that carefully designed curation workflows can approximate gold-standard quality at scale.

At the same time, the experiments highlight the limitations of applying models trained on contemporary English to early modern materials. Performance remains well below the levels typically reported for modern benchmarks, and certain categories of entities, notably Irish personal and place names, continue to pose problems. From a historical perspective, these errors are not distributed randomly. They skew towards the very names and forms that matter most for recovering women's lives in a multilingual, colonial context.

By making these patterns visible, benchmarking reveals the points at which the digital apparatus threatens to reproduce the silencing already present in the archive. It forces us to ask not only

(5) IMPLICATIONS/APPLICATIONS

The VOICES project has used the Funeral Entries as a test bed for exploring how benchmarking can serve as a bridge between data science and historical interpretation. By creating multiple transcriptions of the same manuscript entries, defining a small but carefully annotated benchmark, and systematically comparing NER models and transcription workflows, we have been able to quantify where and how current tools fall short, and to identify workflows that best preserve the visibility of early modern women in the data.

Our findings have several implications for digital humanities and historical research. First, benchmarking is indispensable when working with historical data that diverge significantly from the domains on which standard models are trained. Without explicit evaluation, it is easy to overlook systematic erasures, especially of marginal or non-standard forms such as Irish names or abbreviated women’s names.

Second, benchmark design is itself an interpretive act. Decisions about what counts as a correct entity, how to regularise split or abbreviated forms, and which transcriptions to prioritise all shape the historical picture that emerges. Treating benchmarking as a historical method foregrounds these decisions and invites critical scrutiny of the assumptions embedded in both the tools and the sources.

Third, benchmarking can foster reusable, FAIR-aligned resources for the wider community. The datasets, annotation protocols, and code developed through this work (McShane et al., 2026) are accompanied by an open GitHub repository (<https://github.com/RyFoR/VOICES-FuneralEntries-NER>), enabling others to reproduce, critique, and extend our experiments. In doing so, we hope to contribute not only a specific benchmark for early modern genealogical manuscripts, but also a model for how benchmarking can support transparent, collaborative method-building in the humanities.

Most importantly, this article has argued that benchmarking should be seen not only as a technical stage in a digital workflow, but as a historical method in its own right. By making visible the interactions between archival form, model design, and scholarly interpretation, it helps to ensure that digital tools expand rather than narrow our capacity to hear women’s voices in the early modern archive.

ACKNOWLEDGEMENTS

The authors are grateful to Robin Dresel, Research Intern on the VOICES project, for his invaluable preparation of material in Transkribus. The authors also wish to thank the National Library of Ireland for generously granting access to the manuscript digital files. The Funeral Entries volume GO MS 66 can be accessed in full on the online catalogue of the National Library of Ireland.

The authors acknowledge all those who contributed indirectly to this research, including colleagues who provided methodological insights, archivists and librarians whose expertise supported the broader research process, and the wider VOICES project team for their continued guidance and collaboration.

FUNDING STATEMENT

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. The ADAPT Centre for Digital Content Technology also partially supports this work under the Research Ireland Research Centres Programme (Grant 13/RC/2106_P2). This research was funded in part by the Austrian Science Fund (FWF) [[10.55776/COE12](https://www.fwf.ac.at/en/foyer/grants/10.55776/COE12)].

COMPETING INTERESTS

The authors have no competing interests to declare.

Conceptualisation: BAM, DRY, FVDB; Data curation: DRY, FVDB; Formal Analysis: BAM, DRY; Funding acquisition: JO, DOS; Investigation: BAM, DRY, FVDB, JO, DOS; Methodology: BAM, DRY, FVDB; Software: DRY; Supervision: BAM, DRY, JO, DOS; Validation: FVDB; Visualisation: BAM, FVDB; Writing – original draft: BAM, DRY, FVDB; Writing – review & editing: BAM, JO, DOS.

AUTHOR AFFILIATIONS

Bronagh Ann McShane  orcid.org/0000-0001-7760-621X

School of Histories and Humanities, Trinity College Dublin, Ireland

Diego Rincon-Yanez  orcid.org/0000-0002-8982-1678

WU Vienna University of Economics and Business, Vienna, Austria; Grupo Comunicación Humana, Universidad de Pamplona, Pamplona, Colombia

Felix Vanden Borre  orcid.org/0009-0001-8704-4588

School of Histories and Humanities, Trinity College Dublin, Ireland

Jane Ohlmeyer  orcid.org/0000-0001-9419-7438

School of Histories and Humanities, Trinity College Dublin, Ireland

Declan O’Sullivan  orcid.org/0000-0003-1090-3548

School of Computer Science and Statistics, Trinity College Dublin, Ireland

REFERENCES

- 1641 Depositions Project.** (2010–). *1641 Depositions*. Trinity College Dublin. <https://1641.tcd.ie>
- Debruyne, C., Munnely, G., Kilgallon, L., O’Sullivan, D., & Crooks, P.** (2022). Creating a Knowledge Graph for Ireland’s Lost History: Knowledge Engineering and Curation in the Beyond 2022 Project. *Journal on Computing and Cultural Heritage*, 15(2), 1–25. <https://doi.org/10.1145/3474829>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K.** (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *NAACL HLT 2019–2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies – Proceedings of the Conference*, 1, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
- Mac Curtain, M., & O’Dowd, M.** (Eds.) (1991). *Women in early modern Ireland, 1500–1800*. Edinburgh University Press.
- Mac Curtain, M., O’Dowd, M., et al.** (1992). An agenda for women’s history in Ireland, 1500–1900. *Irish Historical Studies*, 28(109), 1–37. <https://doi.org/10.1017/S0021121400018551>
- McAnlis, V. W.** (2014). *The Consolidated Index to the Records of the Genealogical Office, Dublin, Ireland* (Vols. 1–4). National Library of Ireland. <https://www.nli.ie>
- McShane, B.** (2019). Inter-confessional marriage in early modern Ireland. In S. Ryan (Ed.), *Marriage and the Irish: A miscellany* (pp. 53–56). Wordwell Press.
- McShane, B.** (2025, July). *Hiding in plain sight: Women, agency and Catholicism in the early modern courts* [Keynote lecture]. Catholic Record Society Conference, Hinsley Hall, Leeds, UK.
- McShane, B., & Vanden Borre, F.** (2025, October). *Digitising death: Gender, genealogy, and the experimental recovery of women’s histories in early modern Ireland* [Online seminar presentation]. Digital History Seminar, Institute of Historical Research, London, UK. <https://www.youtube.com/watch?v=0QC0fs8lstM>
- McShane, B. A., Rincon-Yanez, D., Vanden Borre, F., Ohlmeyer, J., & O’Sullivan, D.** (2026). Digitising Death: Benchmarking Genealogical Data and Recovering Women’s Histories in Early Modern Ireland (0.0.2) [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.18174069>
- National Library of Ireland.** (n.d.). *Genealogical Office: Funeral entries*. <https://www.nli.ie>
- Navigli, R.** (2018). Natural language understanding: Instructions for (present and future) use. *IJCAI International Joint Conference on Artificial Intelligence*, 2018-July, 5697–5702. <https://doi.org/10.24963/ijcai.2018/812>
- Nolan, F.** (2020). “The Cat’s Paw”: Helen Arthur, the act of resumption and the Popish pretenders to the forfeited estates in Ireland, 1700–03. *Irish Historical Studies*, 42(162), 225–243. <https://doi.org/10.1017/ihs.2018.31>
- Nolan, F., & McShane, B.** (2022). Introduction: A new agenda for women’s and gender history in Ireland. *Irish Historical Studies*, 46(171), 207–216. <https://doi.org/10.1017/ihs.2022.26>
- O’Dowd, M.** (1999). Women and the Irish Chancery Court in the late sixteenth and early seventeenth centuries. *Irish Historical Studies*, 31(124), 470–487. <https://doi.org/10.1017/S002112140001436X>
- O’Dowd, M.** (2005). *A history of women in Ireland, 1500–1800*. Pearson Longman.
- Ohlmeyer, J.** (2018). *Making Ireland English*. Oxford University Press. <https://doi.org/10.12987/9780300177503>

- Ohlmeyer, J. (2023). *Making empire*. Oxford University Press. <https://doi.org/10.1093/oso/9780192867681.001.0001>
- Palmer, P. (2018). Fugitive identities: Selves, narratives and disregarded lives in early-modern Ireland. In B. Smith & D. Simms (Eds.), *Becoming and belonging in Ireland AD c. 1200–1600: Essays on identity and cultural practice* (pp. 313–327). Cork University Press.
- Piotrowski, M. (2012). *Natural language processing for historical texts* (Synthesis Lectures on Human Language Technologies, 5[2], 1–157). Morgan & Claypool. https://doi.org/10.1007/978-3-031-02146-6_1
- Provatorova, V., Kanoulas, E., et al. (2020). UvA-ILPS at CLEF HIPE 2020: Named entity recognition and linking on historical newspapers. In *CLEF 2020 Working Notes* (CEUR Workshop Proceedings, Vol. 2696). CEUR-WS.
- Qi, P., Zhang, Y., Zhang, Y., Bolton, J., & Manning, C. D. (2020). Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 101–108. <https://doi.org/10.18653/v1/2020.acl-demos.14>
- Randles, A., McKenna, L., Kilgallon, L., Yaman, B., Crooks, P., & O’Sullivan, D. (2024). The Knowledge Graph Explorer for the Virtual Record Treasury of Ireland. *Proceedings of VOILA 2024: The 9th International Workshop on the Visualization and Interaction for Ontologies, Linked Data and Knowledge Graphs Co-Located with the 23rd International Semantic Web Conference (ISWC 2024)*.
- READ-COOP. (2023). *Text Titan I* [Handwritten text recognition model]. In *Transkribus documentation*. <https://readcoop.eu/transkribus>
- READ-COOP. (2025). *Transkribus user documentation*. <https://readcoop.eu/transkribus>
- Rincon-Yanez, D., & Senatore, S. (2022). FAIR Knowledge Graph construction from text, an approach applied to fictional novels. *Proceedings of the 1st International Workshop on Knowledge Graph Generation From Text and the 1st International Workshop on Modular Knowledge Co-Located with 19th Extended Semantic Conference (ESWC 2022)*, 94–108. http://ceur-ws.org/Vol-3184/TEXT2KG_Paper_7.pdf
- Sadek, J., Vlachidis, A., Pickering, V., Humbel, M., Metilli, D., Carine, M., & Nyhan, J. (2024). Leveraging OCR and HTR cloud services towards data mobilisation of historical plant names. *International Journal of Digital Humanities*, 6(3), 237–261. <https://doi.org/10.1007/s42803-024-00091-4>
- Tait, C. (2022). Progress, challenges and opportunities in early modern gender history, c. 1550–1720. *Irish Historical Studies*, 46(170), 244–269. <https://doi.org/10.1017/ihs.2022.43>
- Tual, S., Abadie, N., Chazalon, J., Duménieu, B., & Carlinet, E. (2023). *A benchmark of nested named entity recognition approaches in historical structured documents*. arXiv. <https://doi.org/10.48550/arXiv.2302.10204>
- VARD. (2025). *VARD2: A tool for dealing with spelling variation in historical corpora* [Software]. <https://ucrel.lancs.ac.uk/vard/about/>
- Virtual Record Treasury of Ireland Project. (2022–). *Virtual Record Treasury of Ireland*. <https://virtualtreasury.ie>
- VOICES Project. (2023–). *VOICES: Voices of Women in Early Modern Ireland, 1550–1700*. European Research Council. <https://voicesproject.ie>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, Article 160018. <https://doi.org/10.1038/sdata.2016.18>
- Yaman, B., McKenna, L., Randles, A., Kilgallon, L., Crooks, P., & O’Sullivan, D. (2024). Digital Prosopographical Information in the Virtual Record Treasury of Ireland’s Knowledge Graph for Irish History. *CEUR Workshop Proceedings*, 3724. <https://ceur-ws.org/Vol-3724/paper2.pdf>

McShane et al.
*Journal of Open
 Humanities Data*
 DOI: 10.5334/johd.491

TO CITE THIS ARTICLE:
 McShane, B. A., Rincon-Yanez, D., Vanden Borre, F., Ohlmeyer, J., & O’Sullivan, D. (2026). Digitising Death: Benchmarking Genealogical Data and Recovering Women’s Histories in Early Modern Ireland. *Journal of Open Humanities Data*, 12: 45, pp. 1–13. DOI: <https://doi.org/10.5334/johd.491>

Submitted: 30 November 2025
Accepted: 21 January 2026
Published: 16 March 2026

COPYRIGHT:
 © 2026 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Journal of Open Humanities Data is a peer-reviewed open access journal published by Ubiquity Press.