ONE MINUTE
ONE POSTER
ONE RESEARCH PROBLEM

PhD Symposium 2026 — Proceedings

Cultural Heritage in the Digital Ecosystem

BolDH

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

PhD Programme in Cultural Heritage in the Digital Ecosystem
Department of Classical Philology and Italian Studies
Alma Mater Studiorum-Università di Bologna

HOW TO CITE

Battisti, Tommaso, Enrica Bruno, Mariangela Giglio, Chiara Manca, Martina Pensalfini, and Lorenzo Sabatino, eds. *One Minute, One Poster, One Research Problem: Proceedings of the Cultural Heritage in the Digital Ecosystem PhD Symposium*. Bologna: Università di Bologna, 2026.

ERC: SH5_12 computational modelling and digitisation in the cultural sphere

Last access to the URLs on May 20th, 2026.

Please notify the publisher of any omissions or errors found: **boldh@unibo.it**

The Zenodo community containing the posters for the proceedings is available at https://zenodo.org/communities/chede-phd-symposium-2026.

DOI: https://doi.org/10.6092/unibo/amsacta/8993

EVENT ORGANISATION AND SUPERVISION

Francesca Tomasi

PUBLICATION PROCESS COORDINATION

Enrica Bruno

EDITORS

Tommaso Battisti
Enrica Bruno
Mariangela Giglio
Chiara Manca
Martina Pensalfini
Lorenzo Sabatino

PREFACE

Lorenzo Sabatino

LAYOUT AND GRAPHIC DESIGN

Tommaso Battisti
Chiara Manca

TABLE OF CONTENTS

6 PREFACE

8 TOMMASO BATTISTI
VISUALISING UNCERTAIN AND CONFLICTING CULTURAL HERITAGE DATA

10 ENRICA BRUNO
VISUALISING THE INVISIBLE: FROM THE KNOWLEDGE GRAPH TO THE DIGITAL PLATFORM

12 MARIANGELA GIGLIO
DIGITAL REMAINS.
UNALLOCATED SPACE AS AN ARCHIVAL PROBLEM

14 ARIANNA PASTORINI
TOWARDS A POSSIBLE METHODOLOGY FOR MEASURING DUCTUS IN EARLY MODERN SCRIPTS

16 VALENTINA ROSSETTI
TRACING THE ART MARKET.
DATA, MODELS AND VISUALISATIONS

18 LORENZO SABATINO
BEYOND THE LITERARY MAP

20 CECILIA TAMAGNINI
INTERSECTIONS BETWEEN CULTURAL DOMAINS:
WHICH ONTOLOGIES?

22 REMO GRILLO
NATURAL LANGUAGE AS AN INTERFACE FOR STRUCTURED KNOWLEDGE

24 CHIARA MANCA
EXTRACTING PROVENANCE EVENTS FROM TEI-ENCODED HISTORICAL CORRESPONDENCE

26 STEFANIA MARZOCCHI
DEVELOPING AN OPERATIONAL EXCELLENCE MODEL FOR THE DIGITAL ECOSYSTEM OF A SCIENTIFIC LIBRARY IN A RESEARCH ORGANISATION

28 MARTINA PENSALFINI
OPENGADDA. OPEN DATA

30 GIANMARCO SPINACI
CAN LLMS SPEAK ART?

32 SHIBINGFENG ZANG
TEXT RESTORATION OF HISTORICAL DOCUMENTS

PREFACE

This volume collects the proceedings of the *PhD Symposium 2026*, held on 6 February 2026 within the Doctoral Programme in *Cultural Heritage in the Digital Ecosystem* of the University of Bologna (Department of Classical Philology and Italian Studies, FICLIT), an interdisciplinary PhD programme that integrates digital methodologies with humanistic inquiry to advance the study, preservation, and dissemination of cultural heritage.

The Symposium brought together doctoral candidates from the 39th and 40th cycles and was conceived as a critical Poster Slam, a format designed to foreground clarity, synthesis, and conceptual focus: *one minute, one slide, one research problem to share within the community to create new connections*. Rather than offering comprehensive presentations of ongoing doctoral projects, participants were invited to frame their research as a single, well-defined question. This constraint encouraged a shift from descriptive exposition to epistemological positioning, foregrounding the tensions, uncertainties, and open challenges that characterise contemporary research in Digital Humanities and Cultural Heritage.

All contributions collected in these proceedings reflect the inherently interdisciplinary nature of the PhD programme, spanning a wide range of domains within the humanities and their multifaceted intersections with digital approaches. Across the projects, a number of shared methodological concerns emerge, pointing to a common effort to rethink how cultural heritage data are modelled, interpreted, and communicated in digital environments.

A first major thematic axis concerns knowledge representation and semantic modelling. Several contributions explore the use of ontologies, knowledge graphs, and linked open data to formalise complex cultural phenomena, from literary constraints and archival creators to provenance events and narrative spatiality. These works interrogate not only how knowledge can be structured, but also which epistemological assumptions underpin modelling choices, raising questions about interoperability, granularity, and interpretive transparency. A second axis revolves around information visualisation and the representation of uncertainty. Multiple projects address the challenge of communicating incomplete, ambiguous, or conflicting data without reducing their complexity, advocating for visual approaches that make uncertainty explicit and support critical interpretation. A third cluster of research engages with digital philology and archival practices, particularly in relation to born-digital materials and historical sources. Questions of legitimacy, ethics, and methodology emerge in the analysis of deleted data, the modelling of incomplete or negotiated events in correspondence, and the restoration of damaged texts through language models. Another significant area concerns the integration of Artificial Intelligence and machine learning techniques within cultural heritage research. From multimodal large language models applied to art-historical reasoning, to neuro-symbolic approaches for querying knowledge graphs, and language models for text restoration, these projects explore both the potential and the limitations of AI systems, with particular attention to issues of reliability, interpretability, and evaluation. Finally, several contributions

address the institutional and infrastructural dimensions of Digital Humanities, examining how tools, workflows, and service models are designed, implemented, and sustained within cultural heritage institutions such as libraries. These perspectives emphasise the socio-technical nature of Digital Humanities practices, in which organisational structures, staff expertise, and technological systems are deeply interconnected.

Across all these areas, a common thread can be identified: the centrality of the research question as a methodological device. The Poster Slam format highlights how research in Digital Humanities is shaped not only by tools and data, but also by critical reflection on the epistemological conditions that govern the production, representation, and interpretation of knowledge. The Symposium thus functions not only as a moment of presentation, but as a collective exercise in abstraction and dialogue, fostering connections across projects that might otherwise remain isolated within their specific domains. In this sense, the contributions gathered here do not merely document ongoing research, but outline a shared intellectual landscape in which questions, rather than answers, constitute the primary engine of scholarly inquiry.

The materials presented in these proceedings are organised by academic year, corresponding to the 39th and 40th cycles of the PhD programme, and each contribution is accompanied by up to five keywords to facilitate thematic navigation. Alongside a concise textual rendering of the intervention delivered during the Symposium, each entry provides a reference to a dedicated repository on Zenodo, where the corresponding poster can be accessed. Each poster is assigned a DOI, ensuring persistent identification and enabling citation and reuse within the broader scholarly ecosystem.

We conclude by acknowledging all doctoral candidates who participated in the Symposium for their critical engagement, together with Professor Francesca Tomasi, whose organisation of the event and coordination of the editorial process made this volume possible. Their combined efforts have made this initiative a valuable occasion for exchange within the doctoral community and a meaningful contribution to the broader field of Digital Humanities and Cultural Heritage studies.

Tommaso Battisti, Enrica Bruno, Mariangela Giglio,
Chiara Manca, Martina Pensalfini, Lorenzo Sabatino

E-MAIL tommaso.battisti5@unibo.it ORCID 0009-0002-7980-3334

VISUALISING UNCERTAIN AND CONFLICTING CULTURAL HERITAGE DATA

Tommaso Battisti

KEYWORDS digital humanities; uncertainty; information visualisation

Epistemic uncertainty is a foundational condition of humanities data, often emerging across multiple interrelated dimensions with distinct evidential and hermeneutic implications. Despite the extensive literature on uncertainty visualisation in the Digital Humanities (DH), practical implementations remain scarce, and web-based DH projects rarely treat epistemic uncertainty as a primary organising principle. While the lack of adequate representation risks oversimplifying complex interpretative information and creating an illusion of certainty, representing multiple uncertainty dimensions increases visual and cognitive complexity.

Drawing on the case study based on the data of the Federico Zeri Foundation photographic archive, which catalogues author attributions across a broad corpus of artworks, this contribution presents ongoing doctoral research addressing two questions:

(RQ1) Can epistemic uncertainty function as the primary lens for enquiry and decision-making in cultural heritage visualisations?

(RQ2) Can we use narration as a technique to support the interpretation and disambiguation of multiple epistemic uncertainties in an otherwise solely explorative and visually complex setting?

The dataset offers a valuable context to investigate how epistemic uncertainty can be meaningfully represented and communicated in digital environments. Multiple, often conflicting attributions coexist with heterogeneous evidence and qualifiers expressing varying degrees of doubt, forming debates of different scales. To address this complexity, we adopt a bottom-up, iterative approach based on co-design sessions with domain experts, investigating how to integrate explorative visualisations with narrative layers that guide users through uncertainty and support interpretive engagement. In addressing the cognitive and visual challenges arising from the simultaneous representation of multiple uncertainty dimensions, narration is thus positioned as a mediator between data complexity and informed scholarly decision-making.

This research contributes to ongoing discussions at the intersection of information visualisation and DH, questioning the representational limits of visual solutions for epistemic uncertainty and exploring its potential treatment as a first-class concern in the design of cultural heritage visualisation systems.

POSTER DOI 10.5281/zenodo.19569974

REFERENCES

Battisti, Tommaso, and Marilena Daquino. 2025. 'Uncertainty, Narrativity, and Critical Approaches in Digital Humanities Information Visualisation Projects'. *Proceedings of the 21st Conference on Information and Research Science Connecting to Digital and Library Science*, 1–10.

Drucker, Johanna. 2011. 'Humanities Approaches to Graphical Display'. *Digital Humanities Quarterly* 5 (1). https://www.digitalhumanities.org/dhq/vol/5/1/000091/000091.html.

Drucker, Johanna. 2013. 'Performative Materiality and Theoretical Approaches to Interface'. *Digital Humanities Quarterly 007* (1).

Drucker, Johanna. 2014. *Graphesis: Visual Forms of Knowledge Production*. MetaLABprojects. Harvard University Press.

Mayr, Eva, Paolo Federico, Silvia Miksch, Günther Schreder, Michael Smuc, and Florian Windhager. 2016. 'Visualization of Cultural Heritage Data for Casual Users'. IEEE VIS Workshop on Visualization for the Digital Humanities 1.

Panagiotidou, Georgia, Houda Lamqaddam, Jeroen Poblome, Koenraad Brosens, Katrien Verbert, and Andrew Vande Moere. 2022. 'Communicating Uncertainty in Digital Humanities Visualization Research'. *IEEE Transactions on Visualization and Computer Graphics*, 1–11. https://doi.org/10.1109/TVCG.2022.3209436.

Piotrowski, Michael. 2023. *Uncertainty as Unavoidable Good*. https://doi.org/10.4119/UNIBI/2983506.

Therón Sánchez, Roberto, Alejandro Benito Santos, Rodrigo Santamaría Vicente, and Antonio Losada Gómez. 2019. 'Towards an Uncertainty-Aware Visualization in the Digital Humanities'. *Informatics* 6 (3): 31. https://doi.org/10.3390/informatics6030031.

Windhager, Florian, Paolo Federico, Gunther Schreder, et al. 2019. 'Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges'. *IEEE Transactions on Visualization and Computer Graphics* 25 (6): 2311–30. https://doi.org/10.1109/TVCG.2018.2830759.

Windhager, Florian, Velitchko Andreev Filipov, Saminu Salisu, and Eva Mayr. 2018. 'Visualizing Uncertainty in Cultural Heritage Collections'. In *EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization (EuroRV3)*, edited by Kai Lawonn, Noeska Smit, Lars Linsen, and Robert Kosara. The Eurographics Association. https://doi.org/10.2312/eurorv3.20181142.

Windhager, Florian, Saminu Salisu, and Eva Mayr. 2019. 'Exhibiting Uncertainty: Visualizing Data Quality Indicators for Cultural Collections'. *Informatics* 6 (3): 29. https://doi.org/10.3390/informatics6030029.

Windhager, Florian, Saminu Salisu, Günther Schreder, and Eva Mayr. 2019. 'Uncertainty of What and for Whom-and Does Anyone Care? Propositions for Cultural Collection Visualization'. *4th IEEE Workshop on Visualization for the Digital Humanities (VIS4DH)*, 1–5.

E-MAIL enrica.bruno2@unibo.it ORCID 0009-0007-5620-3834

VISUALISING THE INVISIBLE: FROM THE KNOWLEDGE GRAPH TO THE DIGITAL PLATFORM

Enrica Bruno

KEYWORDS semantic web; knowledge representation; constrained writing; contemporary literature

Every textual work may be seen as the product of regulated combinatorial procedures operating on discrete formal and semantic units. Whether these rules are adopted intentionally, as in the case of constrained writing, or act as implicit conventions, they shape the formal and semantic architecture of the text. Yet current bibliographic cataloguing systems focus almost entirely on the bibliographic product, leaving the generative mechanisms of texts implicit and semantically underspecified. This research addresses this descriptive gap by introducing DeSMòS (Descriptive Semantic Model for Structured Texts), an OWL ontology developed to formally represent literary constraints as generative devices and make them computationally accessible.

The research is driven by two Research Questions (RQs):

RQ1: What is a literary constraint and which entities are necessary to describe it formally?

RQ2: Which strategies for querying and visualising a Knowledge Graph (KG) can make explicit the latent compositional mechanisms structuring literary works, ensuring non-expert accessibility in compliance with copyright restrictions?

To address RQ1, the methodology integrates a domain analysis of potential literature, specifically the production of the Oplepo (Opificio di Letteratura Potenziale), treated as a paradigmatic case study of regulated textual forms, with Semantic Web technologies. DeSMòS is aligned with international standards, including CIDOC CRM, LRMoo, PROV-O, and SKOS, ensuring full interoperability and adherence to FAIR principles. The model formalises three core areas: (i) compositional rules; (ii) creation events, including texts and agents involved; (iii) textual evidence, distinguishing between intratextual traces and authorial declarations.

Building on this formalisation, the research aims to transition from the KG to the design of a digital platform supporting data exploration and graph querying through faceted navigation, visual metaphors, and SPARQL-based interfaces. A central challenge lies in balancing the semantic granularity of structured data with the legal restrictions imposed by copyright. Taken together, the two phases enact a shift in perspective from the description of what a text is to the formal representation of how it is made, opening new analytical possibilities for the study of complex textual forms, including combinatorial structures, linguistic interplay, and hybrid text-image architectures.

POSTER DOI 10.5281/zenodo.19567542

REFERENCES

Andrews, Chris. 2003. «Constraint and Convention: The Formalism of the Oulipo». Neophilologus 87 (2): 223–32. https://doi.org/10.1023/A:1022686129670.

Aragona, Raffaele, a c. di. 2012. L'invenzione e la regola - Il potenziale della letteratura. Edizioni Scientifiche Italiane.

Baetens, Jan. 1997. «Free Writing, Constrained Writing: The Ideology of Form». Poetics Today 18 (1): 1–14. https://doi.org/10.2307/1773230.

Bradley, Adam James, Hrim Mehta, Mark Hancock, e Christopher Collins. 2021. Visualization, Digital Humanities, and the Problem of Instrumentalism. https://hdl.handle.net/10155/1313.

Guarino, Nicola, Daniel Oberle, e Steffen Staab. 2009. «What Is an Ontology?» In Handbook on Ontologies. https://doi.org/10.1007/978-3-540-92673-3_0.

Oulipo. 1973. La Littérature potentielle: créations, re-créations, récréations. Gallimard.

Oplepo. 2005. Biblioteca Oplepiana. Zanichelli.

Possemato, Tiziana. 2024. Entity modeling: la terza generazione della catalogazione. 1a ed. Vol. 11. Biblioteche & bibliotecari / Libraries & librarians. Firenze University Press. https://doi.org/10.36253/979-12-215-0393-7.

Rahaman, Wasim. 2020. «Semantic Web-Linked Data and Libraries». Handbook of Research on Emerging Trends and Technologies in Library and Information Science, 27–42. https://doi.org/10.4018/978-1-7998-8051-6.ch009.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, et al. 2016. «The FAIR Guiding Principles for Scientific Data Management and Stewardship». Scientific Data 3 (1): 160018. https://doi.org/10.1038/sdata.2016.18.

E-MAIL mariangela.giglio2@unibo.it ORCID 0009-0003-2521-1920

DIGITAL REMAINS. UNALLOCATED SPACE AS AN ARCHIVAL PROBLEM

Mariangela Giglio

KEYWORDS archives; born digital; digital philology

The poster addresses unallocated space as an archival problem within born-digital archives. The poster is guided by the following research question: **is it legitimate to analyse what the author had chosen to delete?** While this remains an open and debated issue, the present contribution seeks to address it through the combined lens of forensic methodology, philological analysis, and ethical reflection. Drawing on the case study of the archive of Franco Fortini, it demonstrates how unallocated space may contain digital fragments with textual and genetic relevance, providing access to traces of writing processes no longer present among active files. Although invisible at the level of the file system, such residues are materially preserved in storage media acquired through forensic imaging, thus constituting a liminal zone that exceeds traditional descriptive practices.

From a methodological perspective, the study presents the development of a software tool designed for the analysis and carving of unallocated space, based on the recognition of known file signatures (so-called magic bytes) and on byte-level segmentation of the binary stream. The identification of marker sequences allows the content to be divided into autonomous segments, which are subsequently exported as readable files, thereby making accessible materials that would otherwise remain uninterpretable. The same procedure has also been applied to "blank" media, in order to demonstrate the persistence of information independently of archival or authorial intentionality.

Alongside its technical and epistemological dimensions, the paper raises critical questions concerning the legitimacy of analysing deleted data and the limits of interpretation when forensic evidence exceeds the author's intention. In this context, it argues that the right to knowledge cannot indiscriminately override the right to privacy, making it necessary to develop new descriptive and normative criteria for integrating unallocated space into archival and philological practices.

POSTER DOI 10.5281/zenodo.19571827

REFERENCES

Allegrezza, Stefano. 2020. «Biblioteche e archivi personali in ambiente digitale: le sfide che si profilano all'orizzonte». In Il privilegio della parola scritta: Gestione, conservazione e valorizzazione di carte e libri di persona, a cura di Fiammetta Saba e Giovanni Di Domenico. Associazione italiana biblioteche.

Duranti, Luciana. 2009. «From digital diplomatics to digital records forensics». Archivaria, 39–66.

Duranti, Luciana, e Barbara Endicott-Popovsky. 2010. «Digital Records Forensics: A New Science and Academic Program for Forensic Readiness». Journal of Digital Forensics, Security and Law, pubblicazione online ad accesso anticipato. https://doi.org/10.15394/jdfsl.2010.1075.

Durno, John. 2016. «Digital Archaeology and/or Forensics: Working with Floppy Disks from the 1980s». The Code4Lib Journal, issue 34 (October).

Erway, Ricky. 2010. Defining «Born Digital.» OCLC Research. https://www.oclc.org/content/dam/research/activities/hiddencollections/borndigital.pdf.

Gengenbach, Martin, Alexandra Chassanoff, e Porter Olsen. 2012. «Integrating Digital Forensics into Born-digital Workflows: The BitCurator Project». Proceedings of the American Society for Information Science and Technology 49 (1): 1–4. https://doi.org/10.1002/meet.14504901343.

Gooding, Paul, Jos Smith, e Justine Mann. 2019. «The Forensic Imagination: Interdisciplinary Approaches to Tracing Creativity in Writers' Born-Digital Archives». Archives and Manuscripts 47 (3): 374–90. https://doi.org/10.1080/01576895.2019.1608837.

Kirschenbaum, Matthew G. 2012. Mechanisms: New Media and the Forensic Imagination. First MIT Press paperback edition. The MIT Press.

Kirschenbaum, Matthew G. 2013. «The .txtual Condition: Digital Humanities, Born-Digital Archives, and the Future Literary». Digit. Humanit. Q. 7.

Kirschenbaum, Matthew G. 2019. «La letteratura born digital: gli scrittori e il Word Processor». In Teoria e forme del testo digitale, a cura di Zaccarello Michelangelo. Carocci editore.

Kirschenbaum, Matthew G. 2021. Bitstreams: The Future of Digital Literary Heritage. Material Texts. PENN, University of Pennsylvania press.

Kirschenbaum, Matthew G., Richard Ovenden, Gabriela Redwine, e Rachel Donahue. 2010. Digital Forensics and Born-Digital Content in Cultural Heritage Collections. CLIR Publication 149. Council on Library and Information Resources.

Lebrave, Jean-Louis. 2011. «Computer forensics : la critique génétique et l'écriture numérique». Genesis, issue 33 (October): 137–47. https://doi.org/10.4000/genesis.633.

Ries, Thorsten. 2018. «The rationale of the born-digital dossier génétique: Digital forensics and the writing process: With examples from the Thomas Kling Archive». Digital Scholarship in the Humanities 33 (2): 391–424.

E-MAIL arianna.pastorini2@unibo.it ORCID 0009-0000-1444-5463

TOWARDS A POSSIBLE METHODOLOGY FOR MEASURING DUCTUS IN EARLY MODERN SCRIPTS

Arianna Pastorini

KEYWORDS digital palaeography; women's handwriting; early modern manuscripts

This poster offers a methodological reflection on **whether it is possible to define and measure ductus (understood in the Italian palaeographic tradition as the speed and manner of execution of writing) in quantitative terms (RQ)**. In palaeography, ductus is a key analytical parameter, yet its assessment has traditionally relied on qualitative and subjective judgement, expressed through categories such as "posato", "semi-posato", "semi-corsivo", and "corsivo", which are often applied inconsistently.

The study aims to develop a procedure that complements impressionistic description with a formalised system of measurement. To this end, the research adopts a data-driven approach based on the systematic collection and analysis of observable features. The case study consists of a dataset of 73 graphic samples from the sixteenth century, drawn from the female monastery of Corpus Domini in Bologna.

The methodology focuses on three parameters considered particularly relevant for describing cursivity: the number of external ligatures, the degree of inclination relative to the vertical axis, and the presence or absence of loops or doubled strokes. In line with Noordzij's stroke theory, cursivity is treated as a dynamic property of writing, linked to the continuity of the gesture and the presence of returning movements. The data, collected using a web application developed within the "eManuSkript" project (University of Göttingen), were used to calculate, for each sample, an index of ductus defined as the weighted sum of the normalised values of the three parameters.

Parameter weighting was determined by combining an analysis of data distribution with theoretical considerations. Ligatures were treated as the primary parameter, given their direct relation to the continuity of the writing gesture; inclination was considered a secondary indicator; and loops were used as a corrective feature, particularly to distinguish higher degrees of cursivity from intermediate ones. The weights are currently being refined.

Based on these values, a four-level ductus scale was established. The results show a satisfactory agreement between the algorithmic classification and traditional palaeographic assessment. The model represents a first attempt to quantify ductus, although it remains provisional and partly subjective, as it reflects a specific interpretation of cursivity and relies on a selected set of parameters. Once refined on the present dataset, the model will need to be tested on other corpora of contemporary scripts in order to assess its replicability.

POSTER DOI 10.5281/zenodo.19605051

REFERENCES

Bartoli, Marco. Caterina la santa di Bologna. Bologna: Edizioni Dehoniane Bologna, 2003.

Casamassima, Emanuele. Tradizione corsiva e tradizione libraria della scrittura latina del Medioevo. Roma: Gela Editrice, 1988.

Cherubini, Paolo, and Alessandro Pratesi. Paleografia latina: L'avventura grafica del mondo occidentale. Città del Vaticano: Scuola Vaticana di Paleografia, Diplomatica e Archivistica, 2010.

Derolez, Albert. The Palaeography of Gothic Manuscript Books: From the Twelfth to the Early Sixteenth Century. Cambridge: Cambridge University Press, 2003.

Dorofeeva, Anna, Jeremy Thompson, Mohamed Basuony, Yiyang Huang, Nuray Haskilic, Zhiling Dong, Peter Evans, and Aleyna Yildiz. eManuSkript: Digital Tools for Palaeography. Zenodo, February 20, 2026. https://doi.org/10.5281/zenodo.18999747.

Mastruzzo, Antonino. "Problemi metodologici e prospettive di ricerca nello studio della tradizione grafica corsiva." Litterae Caelestes 1, no. 1 (2005): 29–40.

Noordzij, Gerrit. The Stroke: Theory of Writing. London: Hyphen Press, 2005.

Proietti, Giorgia. "I manoscritti del Corpus Domini di Bologna. Ricognizione e descrizione del fondo conservato nel monastero." Archivum Franciscanum Historicum 116 (2023): 395–486.

E-MAIL valentina.rossetti7@unibo.it ORCID 0009-0007-5385-253X

TRACING THE ART MARKET. DATA, MODELS AND VISUALISATIONS

Valentina Rossetti

KEYWORDS knowledge representation; information visualisation

The Federico Zeri Foundation preserves a collection of particular relevance for the study of the art market, including an extensive photographic archive and approximately 1,900 historical auction catalogues spanning 1869–1929. Collecting and the antiquarian market were central themes throughout Federico Zeri's career: he maintained direct professional relationships with both Italian and international dealers, recognising their crucial role in the artistic landscape and documenting figures on whom bibliography remains scarce. The "Tracing the Art Market" project aims to reconstruct the history of the Italian art market, with a particular focus on antiquarians active between 1860 and 1929 ca., through the collection, analysis, semantic modelling, and online publication of Zeri's data.

The poster addresses a central methodological question: **how can data visualisations be used critically in digital art history, making explicit the criteria of data selection and inference, and representing informational uncertainty in a way that is honest and intelligible for both expert and general audiences?**

The project employs a dual methodology. A data-driven strand collects biographical and relational data on Italian antiquarians from heterogeneous archival sources (i.e. annotated photographs from the Zeri photo library, commercial archives, specialised periodicals, auction catalogues, and testimonies from direct heirs), organising this information in narrative form and in a relational database. In parallel, a data-extraction strand employs a large language model (Mistral-Pixtral) to digitise and extract structured information from the auction catalogues. The resulting data is semantically enriched through an ontological mapping based on CIDOC CRM and the Linked Art application profile, aligned with Getty AAT and ULAN controlled vocabularies. Interoperability is pursued at two levels: metadata (with the Zeri Catalogue and OPAC) and digital assets (with the university digital library AMS Historica).

The envisioned output is a multi-view web interface integrating a browsable catalogue, curated narrative itineraries, and advanced visualisations, such as maps, timelines, and relational graphs. These tools, however, raise critical epistemological concerns regarding the transparency of the interpretive frameworks underlying data selection. Furthermore, they necessitate a rigorous and honest representation of informational uncertainty stemming from incomplete sources, variable digitisation quality, and inferential gaps. The poster presents the current state of the project and its open challenges, contributing to the broader discussion on the epistemologically aware use of visualisations in digital art history.

POSTER DOI 10.5281/zenodo.19632271

REFERENCES

CIDOC Documentation Standards Working Group. CIDOC Conceptual Reference Model (ISO 21127:2023). 2023. https://cidoc.mini.icom.museum/.

Daquino, Marilena, and Francesca Tomasi. "Historical Context Ontology (HiCO): A Conceptual Model for Describing Context Information of Cultural Heritage Objects." In Metadata and Semantics Research: 9th International Conference, MTSR 2015, Proceedings, 424–436. Cham: Springer, 2015. https://doi.org/10.1007/978-3-319-24129-6_37.

Flanders, Julia, and Fotis Jannidis, eds. The Shape of Data in Digital Humanities: Modeling Texts and Text-Based Resources. London: Routledge, 2019.

Fletcher, Pamela, and Anne Helmreich. "Digital Methods and the Study of the Art Market." In The Routledge Companion to Digital Humanities and Art History. New York: Routledge, 2020.

Getty Research Institute. Art & Architecture Thesaurus Online. Los Angeles: Getty Research Institute, 2024. http://www.getty.edu/research/tools/vocabularies/aat/.

Linked Art Community. "Linked Art: A Web Standard for Describing Art, and the Right Way to Describe It." 2024. https://linked.art/.

Livathinos, Nikolaos, Christoph Auer, et al. "Docling: An Efficient Open-Source Toolkit for AI-Driven Document Conversion." arXiv preprint arXiv:2501.17887, 2025. https://doi.org/10.48550/arXiv.2501.17887.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, et al. 2016. «The FAIR Guiding Principles for Scientific Data Management and Stewardship». Scientific Data 3 (1): 160018. https://doi.org/10.1038/sdata.2016.18.

E-MAIL lorenzo.sabatino3@unibo.it ORCID 0009-0007-1496-722X

BEYOND THE LITERARY MAP

Lorenzo Sabatino

KEYWORDS literary cartography; semantic web; spatial humanities; information visualisation

Digital literary maps have become a prominent methodological approach in the field of Spatial Humanities, offering new ways to explore the relationship between narrative space and geographical representation. However, many existing literary mapping projects still largely rely on reductive cartographic logic that translates narrative space into collections of georeferenced points. While this approach works for stable and identifiable locations, it becomes problematic when literary texts operate through ambiguity, relational spatiality, or fictional geography.

This doctoral project proposes to **rethink the epistemological status of literary maps by considering them not as neutral containers of spatial information, but as interpretive devices comparable to texts themselves**. Just like any other text, the map represents a selective and rhetorical cultural artefact: it does not simply represent space but actively structures and mediates knowledge about it. Consequently, the objective of literary cartography should not be to visualise everything that can be mapped, but rather to make explicit the interpretive relationships through which narrative texts construct spatial meaning.

To address this issue, the project integrates approaches from the Semantic Web and Linked Open Data to model the qualitative relationships between narrative space and geographical entities. In this perspective, the modelling of spatial data is conceived as part of a scalable reading approach, in which different levels of abstraction - ranging from textual mentions to more stable spatial constructs - can be dynamically related and analysed. Ontologies and controlled vocabularies are thus employed as epistemic frameworks that organise interpretive categories, preserve distinctions between degrees of referentiality, and enable the representation of spatial ambiguity as a meaningful analytical dimension. Through the use of persistent identifiers (URIs), these relations become machine-readable semantic objects that can be queried, reused, and compared across different projects.

From a visual point of view, this approach leads to hybrid cartographic interfaces in which different spatial ontologies are reflected in distinct visual grammars. Real locations may appear as georeferenced points, while vague areas or narrative spatial constructs can be represented through alternative visual forms that express their epistemic status.

The broader aim of this research is twofold: first, to establish interoperable semantic infrastructures for modelling narrative spatiality, and second, to develop visualisation strategies capable of communicating the interpretive complexity of literary space without reducing it to purely cartographic coordinates. In doing so, the project positions digital literary mapping as a form of scholarly modelling rather than a simple technique of spatial illustration.

POSTER DOI 10.5281/zenodo.19647113

REFERENCES

Bodenhamer, David J., John Corrigan, and Trevor M. Harris, eds. 2010. The Spatial Humanities: GIS and the Future of Humanities Scholarship. Bloomington: Indiana University Press

Bushell, S. 2020. Reading and Mapping Fiction: Spatialising the Literary Text (1ª ed.). Cambridge University Press

Cosgrove, D. E. (ed.). 2002. Mappings (Reprint). Reaktion Books

Eide, Ø. 2015. Media boundaries and conceptual modelling: Between texts and maps. Palgrave Macmillan

Fiorentino, F., Miglio, C., & Valtolina, A. (eds.). 2025. Immaginari cartografici: Riconfigurazioni teoriche dello spazio. Meltemi

Harley, J. B. 1989. Deconstructing the map. Cartographica, 26(2), 1–20.

Jessop, M. 2004. The visualization of spatial data in the humanities. Literary and Linguistic Computing, 19(3), 335–350

Krautter, B. 2023. The scales of (computational) literary studies: Martin Mueller's concept of scalable reading in theory and practice. In F. Armaselu & A. Fickers (Eds.), Zoomland: Exploring scale in digital history and humanities. De Gruyter Oldenbourg

Martin-Rodilla, P., & Gonzalez-Perez, C. (2018). Representing Imprecise and Uncertain Knowledge in Digital Humanities: A Theoretical Framework and ConML Implementation with a Real Case Study. Proceedings of the Sixth International Conference on Technological Ecosystems for Enhancing Multiculturality, 863–871

Piatti, Barbara, Hans Rudolf Bär, Anne-Kathrin Reuschel, Lorenz Hurni, and William Cartwright. 2009. "Mapping Literature: Towards a Geography of Fiction." In Cartography and Art, edited by William Cartwright, 177–92. Berlin: Springer

Pickles JA. 2006. History of Spaces: Cartographic Reason, mapping and the geo-coded world [2004]. London and New York: Routledge

Reuschel, Anne-Kathrin, Barbara Piatti, and Lorenz Hurni. 2013. "Modelling Uncertain Geodata for the Literary Atlas of Europe." In Understanding Different Geographies, edited by Karel Kriz, William Cartwright, and Michaela Kinberger. Berlin: Springer

Stockhammer, R. 2015. Exokeanismós: The (Un) Mappability of Literature

Ungern-Sternberg, A. 2009. Dots, Lines, Areas and Words: Mapping Literature and Narration (With some Remarks on Kate Chopin's "The Awakening"). Lecture Notes in Geoinformation and Cartography

Westphal, B. (2009). Geocritica: Reale, finzione, spazio (Ed. orig. 2007). Armando Editore

E-MAIL cecilia.tamagnini7@unibo.it ORCID 0009-0006-9506-2686

INTERSECTIONS BETWEEN CULTURAL DOMAINS: WHICH ONTOLOGIES?

Cecilia Tamagnini

KEYWORDS archive creators; ontologies; semantic modelling

Over the past decades, starting from 1992 onwards, there have been several attempts to conceive, develop and propose standards for archival description. In particular, it has become increasingly clear that the creating entity of an archive, understood as a "person", requires complex articulation in order to reflect the variety of their history, activities, and the documentation they have produced. The creating entity is not only the producer of an archive, but may also have corresponded with other archival creators, have been an author or the subject of publications, have created a book collection, have been the subject of works of art, or have commissioned or collected them.

The poster addresses two main questions. The first one is: **how can all these aspects be brought together?** The solution is expected: the use of linked open data and ontologies, particularly archival-specific ones such as Records in Contexts (RiC-O), but also those belonging to other domains has increasingly emerged as ideal. The second question is: **How should one choose the ontologies that can take part in this dialogue?**

The selection of RiC-O is substantiated by its status as a formal standard developed by an international commission, specifically the International Council on Archives (ICA), ensuring a globally recognised framework for archival description. Moreover, many international case studies demonstrate that RiC-O can successfully dialogue with other ontologies from different domains.

In the Italian context, in recent years, the ArCo ontology has been proposed: structured into seven interrelated modules, it is based on the General Catalogue of the Italian Ministry of Culture.

The positive aspects that can support the choice of ArCo are its original vocation for networking and its development within the Italian context, thus corresponding to the specific characteristics of the national cultural heritage. On the downside, it is necessary to note the lack of detailed technical documentation and specific elements useful for archival description, which do not allow for in-depth analysis or a detailed representation.

POSTER DOI 10.5281/zenodo.19605717

REFERENCES

Alfier, Alessandro. 2018. “La descrizione archivistica all’interno di sistemi complessi”. In Descrivere gli archivi al tempo di RIC-CM, a cura di Di Marcantonio, Giorgia e Valacchi, Federico, 45-80. Macerata: EUM, 2018

Di Marcantonio, Giorgia. 2018. “Resource Description and Access and the Conceptual Model Records in Contexts. A Conceptual Model for Archival Description: Comparable Objects?”. JLIS.It 9 (1). https://doi.org/10.4403/jlis.it-12412.

Di Marcantonio, Giorgia. 2023. “From Record to Data. New Purposes for Archival Description Processes”. JLIS.It 14 (2). https://doi.org/10.36253/jlis.it-549.

International Council on Archives-ICA. “Expert group on archival description, EGAD”. 2026-03-20. https://www.ica.org/ica-network/expert-groups/egad/records-in-contexts-ric/.

Ministero della Cultura. “dati.cultura.gov.it. Ar.Co.”. 2026-03-20. https://dati.beniculturali.it/arco-rete-ontologie.

Vassallo, Salvatore. 2018. “La descrizione archivistica oggi”. In Descrivere gli archivi al tempo di RIC-CM, a cura di Di Marcantonio, Giorgia e Valacchi, Federico, 31-44. Macerata: EUM, 2018

Vitali, Stefano. 2014 “La descrizione degli archivi nell’epoca degli standard e dei sistemi informatici”. In Archivistica. Teorie, metodi, pratiche, a cura di Giuva, Linda e Guercio, Maria, 179-208. Roma: Carocci, 2014

E-MAIL remo.grillo@unibo.it ORCID 0000-0003-3585-0198

NATURAL LANGUAGE AS AN INTERFACE FOR STRUCTURED KNOWLEDGE

Remo Grillo

KEYWORDS question answering; knowledge graphs; LLM

This PhD project investigates **how answers given by AI systems can become grounded in expert-curated facts with explicit worldviews and provenance**, securing transparent deductive reasoning from knowledge graphs together with the flexibility of natural language, but also the other way around: how access to cultural heritage knowledge graphs can be made easier and more intuitive across domains with the use of Large Language Models. In particular, the research starts from a recurring problem in Knowledge Graph Question Answering: LLMs' performance becomes fragile when applied to big and complex cultural heritage graphs. In this setting, event-centric structures, long multi-hop paths, heterogeneous modelling practices, and the frequent divergence between ontology and instantiated data make query generation particularly error-prone.

The project asks whether a distillation (summarisation) of a knowledge graph can improve NL-to-SPARQL translation according to approaches that rely only on ontology-level descriptions or heavy full-graph injections. More specifically, it explores two research questions: first, whether a hybrid neuro-symbolic architecture can reduce hallucinations; and second, how such an approach can be evaluated in a reproducible way on cultural heritage knowledge graphs.

The project combines graph exploration, symbolic validation, question interpretation, and LLM-based production. A distillation process extracts the "data model" (class-property patterns, multi-hop structures, usage frequencies) from a SPARQL endpoint. These patterns are encoded and injected into prompt-based NL-to-SPARQL generation. The output is validated in layers for syntactic correctness, lexical validity of referenced entities/properties, and compatibility with the actual data model. Detected errors trigger iterative refinement by feeding explicit constraints back into the generation loop.

The research also includes evaluation. The project will create benchmark datasets from cultural heritage graphs, featuring manually curated natural language questions and gold SPARQL queries, to compare zero-shot prompting, ontology-guided, and data-pattern-guided generation. The PhD's main contribution is methodological: defining and testing a framework to ground LLM-based querying in the structure of cultural heritage knowledge graphs, ultimately improving robustness, interpretability, and reuse in Digital Humanities and Cultural Heritage research.

POSTER DOI 10.5281/zenodo.19632673

REFERENCES

Mountantonakis, Michalis, and Yannis Tzitzikas. 2025. "Generating SPARQL Queries over CIDOC-CRM Using a Two-Stage Ontology Path Patterns Method in LLM Prompts." *ACM Journal on Computing and Cultural Heritage* 18 (1): 12:1–12:20. doi:10.1145/3708326.

Varagnolo, Davide, Dora Melo, and Irene Pimenta Rodrigues. 2025. "Translating Natural Language Questions into CIDOC-CRM SPARQL Queries to Access Cultural Heritage Knowledge Bases." *Journal on Computing and Cultural Heritage* 18 (2): 1–28. doi:10.1145/3715156.

Maldonado-Rodríguez, Jose, Arianna Graciotti, Valentina Presutti, and Federico Pianzola. 2025. "Natural Language Querying for Humanities Knowledge Graphs: A Case Study on the GOLEM Knowledge Graph." In *Second International Workshop of Semantic Digital Humanities (SemDH 2025)*. CEUR Workshop Proceedings 4009. CEUR-WS.org.

D'Abramo, Jacopo, Andrea Zugarini, and Paolo Torroni. 2025. "Investigating Large Language Models for Text-to-SPARQL Generation." In Proceedings of the 4th International Workshop on Knowledge-Augmented Methods for Natural Language Processing, 66–80. Association for Computational Linguistics. https://doi.org/10.18653/v1/2025.knowledgenlp-1.5

Sharma, Aditya, Christopher Pal, and Amal Zouaq. 2026. "Reducing Hallucinations in Language Model-based SPARQL Query Generation Using Post-Generation Memory Retrieval." In Findings of the Association for Computational Linguistics: EACL 2026, 4657–4668. Rabat, Morocco: Association for Computational Linguistics. https://doi.org/10.18653/v1/2026.findings-eacl.243

Walter, Sebastian, and Hannah Bast. 2026. "The Wikidata Query Logs Dataset." arXiv preprint arXiv:2602.14594. https://doi.org/10.48550/arXiv.2602.14594

E-MAIL chiara.manca5@unibo.it ORCID 0000-0003-4037-1175

EXTRACTING PROVENANCE EVENTS FROM TEI-ENCODED HISTORICAL CORRESPONDENCE

Chiara Manca

KEYWORDS information extraction; provenance modelling; historical correspondence

Provenance modelling standards are built around completed events: acquisitions that happened, custody transfers that were finalised. This assumption does not hold for source documents that record events as they unfold. Three questions drive this research: **how to extract provenance events from historical correspondence reproducibly**; **how to represent the full event lifecycle, including requests, negotiations, and failed transfers, using existing semantic standards**; and **what gaps those standards present for modelling events before their resolution**.

The case study is the Canneti-Fiacchi correspondence (1711–1730), a 600+ letter exchange between two Camaldolese monks, Pietro Canneti and Mariangelo Fiacchi, documenting the formation of the Biblioteca Classense in Ravenna. Of the full corpus, 133 letters are available in TEI/XML encoding. In this corpus, book transfers are rarely recorded as concluded facts: a letter more often captures a request sent, a price under negotiation, or a deal that may never close. Reconstructing provenance from such sources requires treating incomplete events as primary data, a dimension absent from existing LOD approaches to correspondence and object provenance.

A pilot extraction on 60 letters produced 95 validated provenance events. Two minimum thresholds govern inclusion: items require 2 of 3 criteria (work identification, quantity, physical characteristics); events require 1 of 4 provenance endpoints (origin or destination, agent or location). Events are mapped to a three-layer representational framework: CER-Ontology for epistolary structure, CIDOC CRM for provenance events, and LRMoo for bibliographic objects. Of the 95 events, 36 (38%) are completed, and 59 (62%) are planned or under negotiation; the majority document intermediate phases.

The primary representational gap is temporal: CIDOC CRM provides E8_Acquisition for ownership transfers and E10_Transfer_of_Custody for custody changes, but both classes describe events that have already occurred. Linked Art acknowledges the limit directly: "the model currently cannot describe the future planned event." The proposed approach encodes lifecycle states via P2_has_type within the existing E8/E10 framework. Three further gaps emerged from the corpus (chains of intermediary agents, non-monetary payments such as masses and services, and approximate price qualifiers such as circa and al massimo), for which three minimal ontology extensions are proposed. Whether P2_has_type suffices for lifecycle states or new subclasses are required will be tested when extraction scales to the full 133-letter corpus.

POSTER DOI 10.5281/zenodo.19590576

REFERENCES

Burrows, Toby, Laura Cleaver, Doug Emery, Eero Hyvönen, Mikko Koho, Lynn Ransom, Emma Thomson, and Hanno Wijsman. "Medieval Manuscripts and Their Migrations: Using SPARQL to Investigate the Research Potential of an Aggregated Knowledge Graph." Digital Medievalist 15, no. 1 (2022). https://doi.org/10.16995/dm.8064.

Hyvönen, Eero, Petri Leskinen, and Jouni Tuominen. "LetterSampo — Historical Letters on the Semantic Web: A Framework and Its Application to Publishing and Using Epistolary Data." Journal on Computing and Cultural Heritage 16, no. 1 (2023): 1–23. https://doi.org/10.1145/3569372.

Linked Art. "Promises." Linked Art Data Model. https://linked.art/model/provenance/promises/.

Manca, Chiara, Fiammetta Sabba, Bianca Sorbara, and Silvia Tripodi. "Verso l'edizione digitale del carteggio Canneti-Fiacchi." In Diversità, Equità e Inclusione: Sfide e Opportunità per l'Informatica Umanistica nell'Era dell'Intelligenza Artificiale, Proceedings del XIV Convegno Annuale AIUCD 2025, 184–189. Verona: AIUCD, 2025. http://hdl.handle.net/11585/1019993.

Seibold, Clara, Elena Suárez Cronauer, and Aline Deicke. Correspondence and Epistolary Research Ontology (CER-Ontology), Version 2.0. Academy of Sciences and Literature | Mainz. https://lod.academy/cer/vocab/ontology/.

Shoilee, Sarah Binta Alam, Annastiina Ahola, Heikki Rantala, Eero Hyvönen, Victor de Boer, Jacco van Ossenbruggen, and Susan Legene. "Enhancing Provenance Research with Linked Data: A Visual Approach to Knowledge Discovery." In Proceedings of the Second International Workshop of Semantic Digital Humanities (SemDH 2025), edited by O. Bruns, A. Graciotti, B. Sartini, and T. Tietz, 1–10. CEUR Workshop Proceedings, vol. 4009. Portorož, Slovenia, June 2, 2025. https://ceur-ws.org/Vol-4009/paper_6.pdf.

E-MAIL stefania.marzocchi6@unibo.it ORCID 0000-0001-8777-9720

DEVELOPING AN OPERATIONAL EXCELLENCE MODEL FOR THE DIGITAL ECOSYSTEM OF A SCIENTIFIC LIBRARY IN A RESEARCH ORGANISATION

Stefania Marzocchi

KEYWORDS organisational efficiency; library management; operational excellence in libraries

How do libraries design and manage Digital Humanities (DH) tools, workflows, and service models? How does this approach influence staff capability development, cross-departmental coordination, and operational performance? How can libraries implement people development principles when adopting operational excellence frameworks for DH services, and what systemic enablers or constraints shape staff engagement, learning, and long-term service sustainability?

These research questions examine how libraries apply systemic thinking to the design and management of DH tools, workflows, and service models, with a particular focus on the human dimension of operational excellence. Rather than treating DH services as isolated projects or technical pipelines, the questions frame them as interconnected socio-technical systems in which people, processes, technologies, and institutional structures continuously interact. Within this systemic perspective, staff are not positioned merely as task executors but as critical agents who shape, adapt, and improve operational performance over time.

Central to this inquiry is the extent to which libraries actively involve staff through specialist training, professional development pathways, and cross-functional collaboration in DH services. The questions investigate how such involvement enhances staff autonomy and decision-making capacity, enabling them to identify inefficiencies, reduce waste in workflows (e.g., duplication of effort, bottlenecks, rework), and respond flexibly to changing patrons' needs. By examining how libraries empower staff to understand the "whole system," the research highlights the role of systems literacy in improving coordination across departments such as IT, collection management, digital collection curation, and research support.

Empirically, these questions foreground people as the connective tissue of DH service systems. They emphasise how placing people at the centre, transforming them from simple executors into active protagonists of continuous improvement, contributes to innovation, service quality, and organisational learning. Through workflow mapping, staff interviews, and performance data analysis, the study seeks to demonstrate that operational excellence in library-based DH practices emerges not only from better tools but from empowered staff who can see interdependencies, experiment with improvements, and drive sustainable growth.

POSTER DOI 10.5281/zenodo.19589264

REFERENCES

Carvalho, J. D., and H. Macedo. 2021. "Toyota Inspired Excellence Models." In Learning in the Digital Era, edited by D. J. Powell, E. Alfnes, M. D. Q. Holmemo, and E. Reke, 235–246. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-92934-3_24.

Dulock, M. J., and H. Long. 2015. "Digital Collections Are a Sprint, Not a Marathon: Adapting Scrum Project Management Techniques to Library Digital Initiatives." Information Technology and Libraries 34 (4): 5–17. https://doi.org/10.6017/ital.v34i4.5869.

Holder, S. 2018. "Borrowed from Business: Using Corporate Strategies to Manage Library Projects." In Advances in Library Administration and Organization, vol. 38. Bingley: Emerald Group Publishing Ltd. https://doi.org/10.1108/S0732-067120180000038006.

Ornat, N., and R. Moorefield. 2019. "Library Cartography: Mapping Workflow Processes in an Academic Library." Journal of Library Administration 59 (2): 175–192. https://doi.org/10.1080/01930826.2018.1562805.

Stoddard, M. M., B. Gillis, and P. Cohn. 2019. "Agile Project Management in Libraries: Creating Collaborative, Resilient, Responsive Organizations." Journal of Library Administration 59 (5): 492–511. https://doi.org/10.1080/01930826.2019.1616971.

Yeh, S.-T., M. Arthaud-Day, and M. Turvey-Welch. 2021. "Propagation of Lean Thinking in Academic Libraries." Journal of Academic Librarianship 47 (3): 102357. https://doi.org/10.1016/j.acalib.2021.102357

E-MAIL martina.pensalfini3@unibo.it ORCID 0009-0006-8020-8849

OPENGADDA. OPEN DATA

Martina Pensalfini

KEYWORDS data visualisation; copyright treatment; contemporary literature

How to work around copyright when representing an author and their works?

When engaging with contemporary literature, copyright constraints significantly restrict what researchers may study and represent. This limitation inevitably imposes boundaries on scholarly inquiry into contemporary authors, as was the case with Carlo Emilio Gadda. His works remain under copyright protection, and no agreement has been reached with the estate, effectively precluding the creation of digital facsimiles of his documents.

Various approaches have been adopted to navigate such restrictions, including the use of paywalls to meet heirs' demands, the omission of protected documents from publication, or the provision of metadata records instead of primary materials.

Still, the OpenGadda project was conceived, aiming to establish a digital paradigm that would offer researchers a methodological framework for studying authors whose copyright status remains uncertain or whose documents are not yet publicly accessible.

The project subsequently reoriented its focus away from the author's protected works towards materials and documents in the public domain that are freely available to broad audiences. From these sources, data was extracted through a semi-automated process combining Python-based computational tools with manual verification, ensuring a higher degree of representational accuracy.

The extraction has been applied across the author's archive, personal library, correspondence, and both primary and secondary bibliography. The resulting data was first systematised in structured spreadsheets before being ingested into a back-end platform built on CollectiveAccess, a software suite widely adopted for cataloguing and data management within cultural heritage institutions.

Future development will involve the implementation of a front-end interface through Pawtucket2, CollectiveAccess's companion package, providing users with a stable and navigable environment in which data can be consulted in an interlinked form. The ultimate objective of the project is to construct a scholarly environment that, through the shared organisational axes of time and place, draws connections across the various data sections, beginning from the correspondence, so as to afford the user an insight into the author's intellectual world and creative thought. This will be complemented by a series of data-driven visualisations that articulate thematic pathways and serve as a comprehensive representation of the author's social network and biography.

POSTER DOI 10.5281/zenodo.19626530

REFERENCES

Antonietti, L., E.M. Sanfilippo, and E. Carbé. "Il caso di BiGraFo: una riflessione intorno alle ontologie bibliografiche." *Umanistica Digitale* 9, no. 20 (2025): 441–458. https://doi.org/10.6092/ISSN.2532-8816/21276.

Battershill, C., et al. *Modernist Archives Publishing Project*. 2013. Accessed March 23, 2026. https://www.modernistarchives.com/.

Centro Studi Gadda. *L'archivio degli archivi*. 2024. Accessed March 23, 2026. https://centrostudigadda.unipv.it/images/pdf/20240729_LARCHIVIODEGLIARCHIVI.pdf.

Centre for Manuscript Genetics (University of Antwerp), Beckett International Foundation (University of Reading), Harry Ransom Humanities Research Center (Austin, Texas), and the Estate of Samuel Beckett. *Samuel Beckett Digital Manuscript Project*. 2007–2013. Accessed March 23, 2026. https://www.beckettarchive.org/.

CollectiveAccess. 2024. https://www.collectiveaccess.org/.

Ghelardi, M. *The Burckhardt Source*. 2019. Accessed March 23, 2026. https://burckhardtsource.org/.

Klinowski, M., and K. Szafarowicz. "Digitisation and Sharing of Collections: Museum Practices and Copyright During the COVID-19 Pandemic." *International Journal for the Semiotics of Law – Revue internationale de Sémiotique juridique 36*, no. 5 (2023): 1991–2019. https://doi.org/10.1007/s11196-023-09986-x.

Mississippi Department of Archives and History. *Eudora Welty Digital Archives*. 2024. Accessed November 27, 2025. https://da.mdah.ms.gov/series/welty.

Pereira, E. "Authors' Rights vs. Textual Scholarship: A Portuguese Overview." *Journal of Intellectual Property, Information Technology and Electronic Commerce Law* 14, no. 4 (2023). http://www.jipitec.eu/online-first-articles-1/5788.

Sahle, P. "What Is a Scholarly Digital Edition?" In *Digital Scholarly Editing: Theories and Practices*, edited by M. J. Driscoll and E. Pierazzo, 19–41. Cambridge: Open Book Publishers, 2015. https://doi.org/10.11647/OBP.0095.

E-MAIL gianmarco.spinaci2@unibo.it ORCID 0000-0002-3504-3241

CAN LLMS SPEAK ART?

Gianmarco Spinaci

KEYWORDS AI; iconography; computer vision

This PhD project investigates **what multimodal large language models (LLMs) know about Art History and how they can support art-historical reasoning**. While Digital Art History has traditionally relied on computational methods for large-scale image analysis, recent multimodal LLMs introduce new possibilities for interpreting visual and semantic content. However, their actual understanding of domain-specific knowledge remains unclear.

The first part of the project answers the question: **What do LLMs know about art history?** To answer this question, the project benchmarks three model families: supervised convolutional neural networks, contrastive vision-language models: CLIP, SigLIP, and generative multimodal LLMs: GPT and Gemini, on the classification of Christian saints across three datasets: ArtDL, ICONCLASS, and Wikidata. The evaluation is structured around zero-shot and few-shot settings, testing both label-based classification and contextual enrichment through Iconclass descriptions. Results show that multimodal LLMs consistently outperform other architectures, achieving accuracy above 90% on curated datasets, while performance decreases on noisier collections. Prompt enrichment through textual descriptions generally improves results, whereas few-shot prompting yields inconsistent gains. These findings suggest that recent LLMs encode substantial visual-semantic knowledge but remain sensitive to dataset quality and task design. Another step towards answering the question is Art-Historical Visual Question Answering (VQA), with a benchmark designed with 870 multiple-choice questions created by domain experts. The benchmark is designed to evaluate not only factual knowledge but also interpretative and reasoning abilities, including iconographic identification, stylistic analysis, and contextual understanding. This part is still in development.

Overall, this study provides a dual evaluation framework that moves beyond classification towards reasoning-based assessment. The results demonstrate that multimodal LLMs can effectively support some tasks in Digital Art History, while highlighting the need for grounded knowledge integration and more robust evaluation methods to address domain-specific complexity. Future experiments aim to understand how these models can support art-historical research.

POSTER DOI 10.5281/zenodo.19581863

REFERENCES

Cetinic, E., & She, J. (2022). Understanding and creating art with AI: Review and outlook. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 18(2), 1–22.

Cornelis, B., Dooms, A., Daubechies, I., & Schelkens, P. (2009). Report on Digital Image Processing for Art Historians. In L. Fesquet & B. Torrésani (Eds.), SAMPTA'09, International Conference on Sampling Theory and Applications (p. Special session on sampling and (in)painting). https://hal.science/hal-00452288

Impett, L., & Offert, F. (2023). There Is a Digital Art History (arXiv:2308.07464). arXiv. https://doi.org/10.48550/arXiv.2308.07464

Panofsky, E. (2018). Studies in iconology: Humanistic themes in the art of the Renaissance. Routledge.

Milani, F., & Fraternali, P. (2021). A Dataset and a Convolutional Model for Iconography Classification in Paintings. Journal on Computing and Cultural Heritage, 14(4), 1–18. https://doi.org/10.1145/3458885

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., … Zoph, B. (2024). GPT-4 Technical Report (arXiv:2303.08774). arXiv. https://doi.org/10.48550/arXiv.2303.08774

Posthumus, E. (2020). Iconclass AI test set. Retrieved September, 16, 2024.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., & others. (2021). Learning transferable visual models from natural language supervision. International Conference on Machine Learning, 8748–8763.

Team, G., Anil, R., Borgeaud, S., Alayrac, J.-B., Yu, J., Soricut, R., Schalkwyk, J., Dai, A. M., Hauth, A., Millican, K., & others. (2023). Gemini: A family of highly capable multimodal models. arXiv Preprint arXiv:2312.11805.

Zhai, X., Mustafa, B., Kolesnikov, A., & Beyer, L. (2023). Sigmoid loss for language image pre-training. Proceedings of the IEEE/CVF International Conference on Computer Vision, 11975–11986.

E-MAIL shibingfeng.zhang@unibo.it ORCID 0009-0005-7320-9088

TEXT RESTORATION OF HISTORICAL DOCUMENTS

Shibingfeng Zhang

KEYWORDS language model; text restoration; digital diplomatics

This PhD project investigates the application of pre-trained language models (PLMs) to the automated restoration of Latin diplomatic texts, with a focus on medieval notary documents. The project addresses a significant challenge in historical document studies: the reconstruction of damaged or missing text in low-resource Latin corpora. To this end, the project systematically evaluates a range of PLMs that vary in architecture, training language, and scale, to identify the most effective approach for this specialised restoration task.

The project is structured around the following research questions:

Does adding Ancient Greek and English during pre-training improve performance in Latin text restoration, or is monolingual pre-training exclusively on Latin more effective?

How does the performance of smaller, domain-specific models fine-tuned on Latin compare to large-scale commercial large language models using few-shot prompting in the context of Latin text restoration?

The experimental design distinguishes between two key settings based on whether the length of the missing text is known or unknown, which leads to the evaluation of both encoder-based models and encoder-decoder or decoder-only models. Controlled comparisons between model pairs which share identical architectures but differ in training data allow for a rigorous assessment of the effect of multilingual pre-training on downstream Latin text restoration tasks.

POSTER DOI 10.5281/zenodo.19632819

REFERENCES

Aprosio, A. P., S. Menini, and S. Tonelli (2022). Bertoldo, the historical bert for italian. In Proceedings of the second workshop on language technologies for historical and ancient languages, pp. 68–72.

Assael, Y., T. Sommerschield, A. Cooley, B. Shillingford, J. Pavlopoulos, P. Suresh, B. Herms, J. Grayston, B. Maynard, N. Dietrich, et al. (2025). Contextualizing ancient texts with generative neural networks. Nature 645 (8079), 141–147.

Bamman, D. and P. J. Burns (2020). Latin bert: A contextual language model for classical philology. arXiv preprint arXiv:2009.10053 .

Lazar, K., B. Saret, A. Yehudai, W. Horowitz, N. Wasserman, and G. Stanovsky (2021). Filling the gaps in ancient akkadian texts: a masked language modelling approach. arXiv preprint arXiv:2109.04513 .

Riemenschneider, F. and A. Frank (2023). Exploring large language models for classical philology. arXiv preprint arXiv:2305.13698 .

Roelli, P. (2014). The corpus corporum, a new open latin text repository and tool. Archivum Latinitatis Medii Aevi 72 (1), 289–304.

Strobel, P. B. (2022). Roberta base latin cased v1.

Torres Aguilar, S. (2022, June). Multilingual named entity recognition for medieval charters using stacked embeddings and bert-based models. In R. Sprugnoli and M. Passarotti (Eds.), Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages, Marseille, France, pp. 119–128. European Language Resources Association.