The Past

DARIAH Annual Event 2025

Book of Abstracts

Programme Committee

Agiatis Benardou (Chair)
Georgios Artopoulos (Chair)
Andrea Scharnhorst (Co-Chair)
Kim Ferguson (Co-Chair, BOA Editor)
Edward J. Gray
Nanette Rißler-Pipka
Tibor Kálmán
Tomasz Parkoła
Adeline Joffres
Elena Gigliarelli
Tanja Wissik
Maria Ilvanidou
Carmen Di Meo
Tomasz Umerle

Special thank you to Ana Ester Tavares for assembling the majority of this Book of Abstracts

Links:
DARIAH AE Website Archive
DARIAH AE 2025 Zenodo Community
DARIAH AE 2025: The Past (Video)

DARIAH-EU
Digital Research Infrastructure for the Arts and Humanities

DOI: 10.5281/zenodo.16411471

Table of Contents

Schedule

Date: Tuesday, 17/June/2025 (Internal Day)

11:00am - 11:30am	**Morning coffee break**
11:30am - 1:00pm	**Community Engagement WG Meeting** Session Chair: **Michael Kurzmeier**, DARIAH-EU; michael.kurzmeier@dariah.eu
11:30am - 1:00pm	**#dariahTeach WG Meeting** Session Chair: **Marianne Huang**, Aarhus University; mph@cc.au.dk Session Chair: **Monika Renate Barget**, Maastricht University; m.barget@maastrichtuniversity.nl
1:00pm - 2:00pm	**Lunch break**
2:00pm - 3:30pm	**Digital Humanities Course Registry WG Meeting** Session Chair: **María Goicoechea**, University Complutense of Madrid; mgoico@filol.ucm.es Session Chair: **Anna Woldrich**, Austrian Academy of Sciences (OeAW); Anna.Woldrich@oeaw.ac.at Session Chair: **Iulianna Van der Lek**, CLARIN; i.vanderlek@uu.nl
2:00pm - 3:30pm	**Ethics and Legality in the Digital Arts and Humanities (ELDAH) WG Meeting** Session Chair: **Koraljka Kuzman Šlogar**, Institute of Ethnology and Folklore Research; koraljkak@gmail.com Session Chair: **Walter Scholger**, CLARIAH-AT; walter.scholger@uni-graz.at
2:00pm - 3:30pm	**Library WG Meeting** Session Chair: **Martin Lhotak**, Library of the Czech Academy of Sciences; lhotak@knav.cz Session Chair: **Nanette Rissler-Pipka**, Max Weber Foundation; rissler-pipka@maxweberstiftung.de
2:00pm - 3:30pm	**Multilingual DH WG Meeting** Session Chair: **Alíz Horváth**, Central European University; aliz.horvath06@gmail.com Session Chair: **Paul Joseph Spence**, King's College London; paul.spence@kcl.ac.uk
2:00pm - 5:30pm	**Digital Numismatics WG Meeting** Session Chair: **Rahel C. Ackermann**, Swiss Inventory of Coin Finds; rahel.ackermann@fundmuenzen.ch Session Chair: **David Wigg-Wolf**, Goethe University Frankfurt | Leicester University; wigg-wolf@em.uni-frankfurt.de
2:15pm - 3:45pm	**Text+: Fishbowl discussion “Beyond the Bubble”** Session Chair: **Lukas Weimer**, Göttingen State and University Library; weimer@sub.uni-goettingen.de
3:30pm - 4:00pm	**Afternoon coffee break**
4:00pm - 5:30pm	**Bibliographical Data WG Meeting** Session Chair: **Vojtěch Malínek**, Institute of Czech Literature, Czech Academy of Sciences; malinek@ucl.cas.cz Session Chair: **Róbert Péter**, University of Szeged; robert.peter@ieas-szeged.hu
4:00pm - 5:30pm	**Research Data Management WG Meeting** Session Chair: **Francesco Gelati**, Universität Hamburg; francesco.gelati@uni-hamburg.de Session Chair: **Françoise Gouzi**, DARIAH Open Science Officer; francoise.gouzi@dariah.eu
5:45pm - 8:45pm	**Welcome Evening & DARIAH WG Showcase** Session Chair: **Kim Ferguson**, DANS; kim.ferguson@dans.knaw.nl Session Chair: **Nanette Rissler-Pipka**, Max Weber Foundation; rissler-pipka@maxweberstiftung.de

Date: Wednesday, 18/June/2025 (Day One of Annual Event)

8:30am - 9:30am	**Welcome coffee**
9:30am - 11:00am	**Plenary | Opening & DARIAH-DE Showcase** Location: **Adam-Von-Trott Saal (Alte Mensa venue)** Session Chair: **Nanette Rissler-Pipka**, Max Weber Foundation; rissler-pipka@maxweberstiftung.de Session Chair: **Daniel Kurzawe**, SUB Göttingen / University of Göttingen; kurzawe@sub.uni-goettingen.de Session Chair: **Stefan Buddenbohm**, Göttingen State and University Library; buddenbohm@sub.uni-goettingen.de
11:00am - 11:30am	**Morning coffee break**
11:30am - 1:00pm	**Panel | How to provide and use bibliographical data for research – the example of the VD17** Session Chair: **Nanette Rissler-Pipka**, Max Weber Foundation; rissler-pipka@maxweberstiftung.de Session Chair: **Thea Lindquist**, University of Colorado; thea.lindquist@colorado.edu
11:30am - 1:00pm	**Topic: Digital Storytelling** Session Chair: **Tomasz Umerle**, PCSS (Poznan Supercomputing and Networking Center); DARIAH-PL; tomasz.umerle@ibl.waw.pl
1:00pm - 2:00pm	**Lunch break**
2:00pm - 3:30pm	**Panel | Archiving and Preservation of Oral History: Recorded Voices of the Past in the Digital Age** Session Chair: **Andrea Scharnhorst**, Data Archiving and Networked Services, Royal Netherlands Academy of Arts and Science; andrea.scharnhorst@dans.knaw.nl Session Chair: **Jetze Jacob Touber**, DANS-KNAW; jetze.touber@dans.knaw.nl
2:00pm - 3:30pm	**Topic: Interdisciplinary Approaches** Session Chair: **Tanja Wissik**, Austrian Academy of Sciences; tanja.wissik@oeaw.ac.at
3:30pm - 4:00pm	**Afternoon coffee break**
4:00pm - 5:30pm	**Keynote | Gabo Arora, "Bicycle of the Heart"** Session Chair: **Georgios Artopoulos**, The Cyprus Institute; g.artopoulos@cyi.ac.cy Session Chair: **Agiatis Benardou**, DARIAH-EU; a.benardou@dcu.gr
6:15pm - 7:30pm	**JRC Meeting** Session Chair: **Andrea Scharnhorst**, Data Archiving and Networked Services, Royal Netherlands Academy of Arts and Science; andrea.scharnhorst@dans.knaw.nl

Date: Thursday, 19/June/2025 (Day Two of Annual Event)

8:30am - 9:30am	**Welcome coffee**
9:30am - 11:00am	**Demonstration Session** Session Chair: **Kim Ferguson**, DANS; kim.ferguson@dans.knaw.nl
9:30am - 11:00am	**Poster Session** Session Chair: **Alexander Steckel**, Göttingen State and University Library; steckel@sub.uni-goettingen.de Session Chair: **Stefan Buddenbohm**, Göttingen State and University Library; buddenbohm@sub.uni-goettingen.de
11:00am - 11:30am	**Morning coffee break**
11:30am - 1:00pm	**Panel | MediaWiki-based tools and services in Digital Humanities workflows** Session Chair: **David Lindemann**, UPV/EHU University of the Basque Country; david.lindemann@ehu.eus
11:30am - 1:00pm	**Topic: Reconstructed Histories** Session Chair: **Edward J. Gray**, CNRS; edward.gray523@gmail.com
12:00pm - 12:30pm	**Extended Demo Session: DARIAH-Campus Content Management System** Session Chair: **Kim Ferguson**, DANS; kim.ferguson@dans.knaw.nl Session Chair: **Vicky Garnett**, DARIAH-EU; vicky.garnett@dariah.eu Extended Demonstration Session from Vicky Garnett, "Introducing the new DARIAH-Campus Content Management System"
1:00pm - 2:00pm	**Lunch break**
2:00pm - 3:30pm	**Topic: LLMs in Action** Session Chair: **Maria Ilvanidou**, Digital Curation Unit, IMSI, Athena RC; m.ilvanidou@dcu.gr
2:00pm - 3:30pm	**Topic: Let's Talk Infrastructure** Session Chair: **Tomasz Parkoła**, Poznan Supercomputing and Networking Center; tparkola@man.poznan.pl
3:30pm - 4:00pm	**Afternoon coffee break**
4:00pm - 5:30pm	**Plenary | DARIAH WG Funding Showcase & more** Session Chair: **Andrea Scharnhorst**, Data Archiving and Networked Services, Royal Netherlands Academy of Arts and Science; andrea.scharnhorst@dans.knaw.nl Session Chair: **Agiatis Benardou**, DARIAH-EU; a.benardou@dcu.gr
7:00pm - 10:00pm	**Social Dinner in Göttingen**

Date: Friday, 20/June/2025 (Day Three of Annual Event)

8:30am - 9:30am	**Welcome coffee**
9:30am - 11:00am	**Topic: Transforming Digital Methods** Session Chair: **Nanette Rissler-Pipka**, Max Weber Foundation; rissler-pipka@maxweberstiftung.de
9:30am - 11:00am	**Topic: Databases, from Past to Future** Session Chair: **Amelia McConville**, DARIAH-EU; amelia.mcconville@dariah.eu
11:00am - 11:30am	**Morning coffee break**
11:30am - 12:30pm	**Plenary | Finale** Session Chair: **Agiatis Benardou**, DARIAH-EU; a.benardou@dcu.gr Session Chair: **Georgios Artopoulos**, The Cyprus Institute; g.artopoulos@cyi.ac.cy
12:30pm - 1:00pm	**Plenary | Closing Remarks** Session Chair: **Georgios Artopoulos**, The Cyprus Institute; g.artopoulos@cyi.ac.cy Session Chair: **Agiatis Benardou**, DARIAH-EU; a.benardou@dcu.gr
2:00pm - 6:00pm	**Women Writers in History WG Meeting** Session Chair: **Alicia Montoya**, Radboud University; alicia.montoya@ru.nl Session Chair: **Amelia Sanz**, Complutense University of Madrid; amsanz@ucm.es Session Chair: **Nina Geerdink**, Utrecht University; n.geerdink@uu.nl

Keynote

Gabo Arora: Bicycle of the Heart

Time: Wednesday, 18/June/2025: 4:00pm - 5:30pm
Session Chair: **Georgios Artopoulos,** The Cyprus Institute
Session Chair: **Agiatis Benardou**, DARIAH-EU

Gabo Arora is a world renowned multi-award winning immersive artist, professor, entrepreneur and former UN diplomat who works with the most cutting-edge emerging technologies, including virtual and augmented reality, to tell some of the most important stories of our time. Widely recognized as a pioneer of new documentary formats, his work, part of the permanent collection of the Museum of Modern Art in New York (MoMA), has been described by the BBC and LA Times, amongst many others, as “game changing”, “powerful, moving and without precedent”, and “transcending all the typical barriers of rectangular cinema.”

“Bicycle of The Heart”: In 1984, when asked to describe what personal computers are good for, Steve Jobs famously called them a “bicycle of the mind”. Now, almost forty years later, spatial computing is here and the totality of its effects will be far more profound; bringing into question our very existence, and what it means to be conscious, alive and human. What if instead of the mind we aimed to create a bicycle – for the heart? Where the power of these immersive technologies are harnessed to connect and enlighten us rather than distract and divide; where we become more than just passive consumers and into active creators and change makers of the future we need. It is possible and it is happening. This keynote charts the journey forward through case studies and a framework for action.

Papers and Panels

Panel | How to provide and use bibliographical data for research – the example of the VD17

Time: Wednesday, 18/June/2025: 11:30am - 1:00pm
Session Chair: **Nanette Rissler-Pipka**, Max Weber Foundation
Session Chair: **Thea Lindquist**, University of Colorado

How to provide and use bibliographical data for research – the example of the VD17

Nanette Rissler-Pipka[1], Thea Lindquist[2], Eetu Mäkelä[3], Hartmut Beyer[4], Maximilian Görmar[4], Peter Kiraly[5], Saskia Limbach[6], Michaela Scheibe[7], Karin Schmidgall[8]

[1]Max Weber Foundation, DARIAH-DE, Germany; [2]Center for Research Data and Digital Scholarship, University of Colorado Boulder, USA; [3]Department of Digital Humanities, University of Helsinki, Finland; [4]Herzog August Bibliothek Wolfenbüttel, Germany; [5]GWDG, Gesellschaft für wissenschaftliche Datenverarbeitung Göttingen, Germany; [6]Georg-August Universität Göttingen, Germany; [7]Staatsbibliothek zu Berlin, Preußischer Kulturbesitz, Germany; [8]Literaturarchiv Marbach, Germany; rissler-pipka@maxweberstiftung.de, thea.lindquist@Colorado.EDU, eetu.makela@helsinki.fi, beyer@hab.de, goermar@hab.de, peter.kiraly@gwdg.de, saskia.limbach@theologie.uni-goettingen.de, Michaela.Scheibe@sbb.spk-berlin.de, Karin.Schmidgall@dla-marbach.de

This panel aims to engage an international audience to discuss the past, present and future of a key resource in European book history and cultural history: Union Catalog of Books Printed in German-Speaking Countries in the 17th Century (VD17). It is assembled, curated and provided by nearly 50 libraries.

The VD17 provides an excellent example of a high-quality retrospective national bibliographical database of the sort that many researchers would like to use as research data. German researchers (Lauer et al. 2024) recently discussed the possibilities and needs regarding this valuable resource and their promise as (FAIR) research data.

Drawing on the DARIAH Bibliodata Working Group's report (Umerle et al. 2022) addressing the joint agendas of stakeholders in the bibliographical data landscape in the humanities as a framework, the panel participants will discuss:

1. The infrastructure needed for the creation, provision and curation of bibliographic metadata,
2. Concrete examples of bibliodata analysis; its challenges and research outputs,
3. Forms of collaboration between libraries, researchers and infrastructure providers regarding the exchange of data and knowledge.

Panel contributions will be 5-minute-pitches, so that the discussion becomes the most important part. Chairs and more members of the DARIAH WG bibliodata intend to join the discussion.

1. The VD17 in the bibliographical data infrastructure landscape

Hartmut Beyer: *VD17: The bibliography of 17th-century German books as part of the national research data infrastructure*
While the VD17 was conceived as an instrument for humanities research, there is still a lot to be done for its usability. The further development aims at a common working environment for all three VDs, the integration of digitised material and full texts as well as the provision of reusable research data.

1. Data analysis and beyond

Eetu Mäkelä, Thea Lindquist: *Analyzing the publications of members of an early modern German academy (or, the sword cuts both ways) - opportunities and challenges presented by the VD17*

The 17th-century German academy Fruchtbringende Gesellschaft was the first and largest of its kind and has attracted much scholarly interest. The VD17 offers an unparalleled opportunity to interrogate the publications with which its many members were associated.

Maximilian Görmar: *The Republic of Letters beyond letters - Analyzing scholarly networks with VD17 data*
With the use of digital methods, the view of the Republic of Letters as an epistolary network became more pronounced in recent years (Hotson/Wallnig 2019). The VD17 data offer opportunities to give new insights beyond that approach (Görmar 2024). The limitations of the VD17 as well as its opportunities for research shall be discussed in the contribution.

Saskia Limbach, Michaela Scheibe: *Working with VD17 - restrictions and opportunities*

There are advantages and disadvantages of some VD17 features, such as download options for larger corpora of data, and the focus on a single century. We will highlight how a focus on the early modern period and including additional data offers more possibilities for researchers in the future.

1. Bibliodata as a collaborative effort?

Peter Kiraly: *Library catalogue as research data - some problems and suggestions*

There has been a growing demand for bibliographic data as research data, and in parallel, libraries are publishing their catalogues. However, the analysis of bibliographic data also requires the researcher to overcome problems (data not available, data structure specially coded, semantic units of library data not evident) which will be discussed and solutions offered.

Karin Schmidgall: *A different look at catalog data - constant or variable?*

Curating catalogue data is an ongoing task for libraries. There are several topics to be considered: What factors influence the data quality of catalogues? How can we build workflows that make it possible to feed catalogue data enriched in projects back into the library catalogues in order to sustainably increase data quality?

Topic: Digital Storytelling

Time: Wednesday, 18/June/2025: 11:30am - 1:00pm
Session Chair: **Tomasz Umerle**, PCSS (Poznan Supercomputing and Networking Center); DARIAH-PL

Merchants of Istanbul: A Visual Novel for Teaching the Early Modern Ottoman Balkans through Interactive Digital Storytelling

Ninja Bumann, Stephanie Lotzow, Sina Roggenkamp
Justus Liebig University Giessen, Germany; ninja.bumann@geschichte.uni-giessen.de, stephanie.lotzow@germanistik.uni-giessen.de, sina.roggenkamp@admin.uni-giessen.de

This paper presents *Merchants of Istanbul*, a serious game in the form of a visual novel designed to teach students basic knowledge about the early modern Ottoman Balkans. The game immerses players in the historical journey of a caravan travelling through Southeast Europe in the late 16th century, bringing to life the geopolitical dynamics, trade routes, and daily life of the various ethnic, linguistic and religious groups within the Ottoman Empire.

Based on historical textual sources in various languages and visual material, *Merchants of Istanbul* offers students the opportunity to experience semi-fictional adventures of a caravan travelling through the early modern Ottoman Balkans, and more precisely from Lemberg/Lviv (in present-day Ukraine) to Istanbul via various towns in present-day Romania and Bulgaria. The main objective of the game is to become wealthy by purchasing luxury goods and safely bring them to Istanbul. By doing so, students acquire knowledge about the society, its ethnic, religious, and linguistic background, government structures as well as trade practices in the Ottoman Empire through an engaging narrative. Besides text and dialogue, which are the main features of a visual novel, the game contains various gameplay features to promote active learning: These include navigable world and town maps, time-sensitive mini-games (quick time events), a travel journal to track progress, an interactive card game to negotiate resources for the best discount, and a source-based glossary that allows students to deepen their knowledge on the history and historiography of the Ottoman Empire. By playing the game and studying the glossary, *Merchants of Istanbul* aims at teaching students to develop critical skills in historical inquiry.

Through gameplay, *Merchants of Istanbul* makes the history of the early modern Ottoman Balkans more accessible to students, especially in Western Europe where this is typically not part of the schools' curriculum. Additionally, the paper illustrates the importance and challenges of interdisciplinary approaches in serious game design: By combining knowledge from historians and game designers from various academic backgrounds as well as programmers and illustrators, *Merchants of Istanbul* is the result of joint academic communication that aims at promoting ludic teaching in the field of digital humanities.

Transforming Cultural Heritage with Extended Reality: Insights from the HERIFORGE Project

Maciej Glowiak, Tomasz Parkola, Michal Kosiedowski, Mikolaj Wegrzynowski
Poznan Supercomputing and Networking Center, Poland; tparkola@man.poznan.pl, mwegrzyn@man.poznan.pl

Over the past twenty years, the approach to cultural heritage has evolved from a focus on preservation to one centered on its broader value. This shift recognizes the importance of cultural heritage across various sectors, where advanced technologies are crucial in addressing its challenges. The HERIFORGE project builds on this idea by connecting ecosystems in Poland, Cyprus, and Turkey -countries where cultural heritage is a core national value. The project aims to foster innovation in both cultural heritage and creative industries, utilizing extended reality (XR) technologies and digital cultural assets to strengthen social resilience. HERIFORGE brings together academic institutions, SMEs, public authorities, and community actors, following the quadruple helix model. The three hubs collaborate to develop a long-term vision (so called strategic impact package) for integrating XR technologies and utilizing digitized cultural heritage assets stored in data spaces and cloud platforms, ensuring their reuse and contribution to future initiatives. Ultimately, the three hubs are to create a pan-regional HERIFORGE Hubs Network for innovation in cultural heritage and XR.

As part of this broader vision, the HERIFORGE project includes the Joint Research Pilot Projects (JPRPs), which focus on exploring the practical applications of XR technologies in cultural heritage. These pilots are designed to address key challenges, such as digitisation, data management, and social inclusion through the use of Virtual Worlds. By creating federated repositories and providing quality CH datasets, the pilots will enhance XR applications in sectors like culture and tourism, fostering innovation and new business opportunities. The pilots also seek to revive lost cultural heritage, offering displaced individuals a sense of belonging by engaging them with their cultural memories in immersive XR environments. Furthermore, the pilots will explore the use of gamification and storytelling techniques in Virtual Worlds to convey historical narratives and transfer knowledge, focusing on enhancing well-being and positive social impacts. JPRPs are led by the hubs in Cyprus, Poland, and Turkey, each contributing expertise in areas such as digitisation, storytelling, gamification, and data orchestration. The collaborative efforts across these hubs will create a unified platform for sharing and accessing digital assets, promoting greater inclusion and cultural understanding. Ultimately, the pilots will demonstrate how XR technologies can play a key role in bridging cultural gaps, revitalising heritage, and fostering social resilience.

As an extension to JPRPs, the HERIFORGE project will organize open calls to identify and fund innovative third-party projects that address specific challenges in using immersive technologies for cultural heritage and social resilience. These calls aim to stimulate innovation, especially in culture, tourism, and social inclusion, while fostering new business opportunities for SMEs. By selecting approx. 12 projects, HERIFORGE will support the development of technical solutions that align with the project's research and innovation challenges.

The HERIFORGE will also foster engagement with international organisations and initiatives, data spaces, other excellence hubs as well as established European research infrastructures – especially DARIAH and social sciences and art communities.

Greece's Difficult Past in Virtual Reality: Commemorating Block 15 Through Digital Immersion

Agiatis Benardou
DARIAH ERIC; agiatis.benardou@dariah.eu

Since the 1970s, the commemoration and preservation of 'difficult heritage,' a term coined by Sharon Macdonald over fifteen years ago, has become a subject of increasing public attention. In the escalation of the European historical turn to memory, we are witnessing the emergence of a new dimension: the distinction of place through reference to historical narrative, whereby historical content is legitimized through exhibitions, memorial plaques, and other modes of urban commemoration. However, despite the opportunities afforded by immersion, there has been a lack of substantive evidence to evaluate current approaches and guide future developments, especially in difficult heritage sites. Particularly in Europe, immersion has not been employed widely in such sites.

This talk will discuss and expand on the affordances and challenges of designing, developing, and assessing the first Virtual Reality production in Greece on Block 15, co-funded by the Greek-German Fund for the Future and the Greek Ministry of Culture.

Block 15 was the building that served as an isolation and torture space within the Concentration Camp of Haidari, Attica, Greece, during 1943 and 1944.

Furthermore, the talk will build on the theoretical and applied approaches to the design and employment of immersive technologies to reconstruct difficult pasts at heritage sites around the world, as discussed in the volume *Difficult Heritage and Immersive Experiences* (1st ed.). Routledge (2022), which was co-edited by "Block 15" project members, Drs. Agiatis Benardou and Anna Maria Droumpouki.

"Block 15" aimed at identifying and re-purposing archival and historical resources toward the development of an immersive VR production on the tangible and intangible heritage of the site. To that end, a series of challenges had to be addressed and overcome, ranging from the overarching methodology, the point of view and narrative backbone of the digital storytelling, the development of historically accurate assets, and the integration of findings from user experience surveys carried out for the purposes of the production.

In addition to the talk, a 2D version (video) of the final production, which was submitted in January 2025, will also be presented, offering an alternative means of engaging with the project's outcomes and reflecting on the potential of immersive media in difficult heritage interpretation.

Spatially-distributed narratives: Generative Ambiguity in Heritage Visualisation

Colter Eugene Wehmeier[1], Georgios Artopoulos[2]
[1]University of Illinois Urbana Champaign, United States of America and The Cyprus Institute, Cyprus; [2]The Cyprus Institute, Cyprus; wehmeie2@illinois.edu

Interactive visualization and virtual reconstruction are reshaping museum engagement by fostering participatory knowledge elicitation and crowdsourcing. While photorealistic digital heritage approaches prioritize accuracy (Parker and Saker 2020; Petrelli 2019), this paper argues that such practices risk limiting deeper cognitive engagement, particularly when the goal is dialogic engagement rather than dissemination. Drawing on participatory heritage and digital humanities scholarship, the study proposes that strategically designed ambiguity in visual representations can instead invite visitors to actively interpret and co-construct meaning, aligning virtual reconstructions with the intellectual and cultural aims of the Digital Humanities and decolonized cultural values.

This research examines modern architectural heritage intertwined with contested histories through Nicosia International Airport (1968), abandoned in Cyprus's UN buffer zone since 1974. The site's significance lies in its layered narratives—as a symbol of modernization, a time capsule of de Certeau's (1984) 'spatial practices' of the everyday (e.g., public terraces and amenities), and a monument to abrupt historical rupture. Its virtual reconstruction confronts the urgency of integrating living memory and embodied knowledge with archival documentation, particularly as firsthand experiences of its social role fade.

To address this, a year-long museum installation featured an interactive virtual environment of the airport, set in 1969—before the division of the country—and intentionally embedding spatial, temporal, and sensory interpretation gaps. These gaps transformed the visualization into a collaborative research instrument and conduit for collective storytelling, inviting users to bridge omissions through personal memories, historical hypotheses, and creative inputs. The approach prioritized democratizing access to cultural heritage and positioned the virtual model as a dynamic, community-informed work-in-progress.

Two methodological frameworks structured the experience:

1. **Scaffolded Interactions through Meaningful Play** (Salen and Zimmerman 2003): Context-dependent design balanced accuracy and ambiguity. Foundational elements (e.g., architectural geometry, historical timelines) were rendered precisely, while speculative or socially charged spaces (e.g., public terraces central to pre-conflict daily life) employed abstract or incomplete visualization. This duality encouraged visitors to negotiate interpretations through speculative scenarios and creative dialogue.
2. **Feedback Mechanisms**: Real-time annotation tools allowed visitors to critique ambiguities, share narratives, and propose revisions, fostering a sense of ownership and sustaining the reconstruction as a reflective, evolving discourse. These mechanisms enabled crowdsourcing knowledge and ensured the model remained a platform for responsible reflective discourse.

The installation demonstrated that strategic ambiguity, when coupled with structured interactivity, deepened engagement by transforming passive observation into active inquiry. Visitors collaboratively debated the airport's cultural meanings, contributing perspectives often excluded from institutional narratives. Critically, the balance between accuracy and abstraction proved foundational: overly ambiguous representations confused users, while excessive realism discouraged creative participation.

By privileging abstraction over photorealism, this approach redefines virtual reconstructions as dynamic spaces for negotiating memory, identity, and loss. The study advances a framework for heritage visualization where designed ambiguity becomes a deliberate tool for social reflection, urging practitioners to embrace incompleteness as a catalyst for participatory meaning-

making. This methodology aligns with decolonized cultural values, positioning digital heritage not as a static replica but as a platform for communities to articulate evolving relationships to place and history.

Panel | Archiving and Preservation of Oral History: Recorded Voices of the Past in the Digital Age

Time: Wednesday, 18/June/2025: 2:00pm - 3:30pm
Session Chair: **Andrea Scharnhorst**, Data Archiving and Networked Services, Royal Netherlands Academy of Arts and Science
Session Chair: **Jetze Jacob Touber**, DANS-KNAW

Archiving and Preservation of Oral History: Recorded Voices of the Past in the Digital Age

Jetze Jacob Touber[1,6], Sanneke Stigter[2], Marijn Braam[3], Annette Langedijk[4], Maarten Heerlien[5], Norah Karrouche[5,6]
[1]DANS-KNAW; [2]University of Amsterdam; [3]Oral History hub 'Sprekende geschiedenis'; [4]SURF; [5]Vrije Universiteit Amsterdam; [6]CLARIAH-NL; jetze.touber@dans.knaw.nl, S.Stigter@uva.nl, marijn@sprekendegeschiedenis.nl, annette.langedijk@surf.nl, n.f.f.karrouche@vu.nl

A panel arranged under the aegis of CLARIAH-NL and SSHOC-NL

Oral History holds great potential to unlock the experiences and knowledge of the past, and to collect perspectives which do not emerge from traditional historical sources, such as archival records and printed sources. Oral History, dealing with digital audio and video content, has also enthusiastically embraced the digital methods and infrastructure developed in the past two decades. Online archiving facilities, increased computing performance, and applications for annotating and analysing audio and video content hold tremendous potential to boost the use and re-use of interview recordings. With the advent of AI, yet a new phase in valorising Oral History materials has already begun. With all these developments, however, also come concerns regarding the long-term management, the accessibility and the responsible use of Oral History sources.

In the span of a couple of years, several projects in the Netherlands have started to look into these various aspects of Oral History in the digital age. These projects aim to develop generic research infrastructure for the management and analysis of Oral History data, but also to reflect critically on their use and to give guidance to researchers. They contribute to the following developments:

- Infrastructure for archiving and publishing Oral History data is being built to support long term preservation and re-use in a legally sound way.
- Workflows and standards (metadata and files) are being identified for making Oral History data findable, accessible, interoperable and reusable.
- AI-tools are being experimented with for enriching Oral History data with automated transcripts, subtitles and topical keywords.
- An ethical code is being developed, for academic researchers, heritage professionals and community archives to work with Oral History data and supporting facilities in a way that does not harm the interviewees or the communities to which they belong.

Partners which collaborate in these projects comprise a wide range of academic institutes, heritage organizations and infrastructure providers operating both at a local and at a national level, including CLARIAH-NL and ODISSEI (the national infrastructures for the Humanities and Social Sciences, respectively). The goals of these Oral History projects complement each other. Together they should provide the variegated user groups feeding into the Oral History community with firm footing to collect data and conduct research in an advanced and responsible manner.

In this panel, representatives of these Oral History-oriented projects will present the various perspectives mentioned above:

- Sanneke Stigter (UvA) and Jetze Touber (DANS-KNAW): "Developing infrastructure for Oral History data archiving and reuse: the OH-SMArt, OH-CORE, SSHOC-NL projects"
- Marijn Braam (Oral History Hub): "The Hidden Stories Project: Mapping Bottlenecks in Digital Infrastructure with Local Oral History Initiatives to Increase Accessibility and Reusability"
- Annette Langedijk (SURF): "Meaningful Memories: AI-Powered Annotation for Discovering and Connecting Concepts in Interviews"
- Maarten Heerlien (VU) and Norah Karrouche (VU): "Drafting an Ethical Code for the archiving and reuse of Oral History data: the StoRe project"

Together, these presentations will showcase how the complementarity of the various Oral History projects advance responsible practices of managing, sharing and analysing interview materials in a careful manner. They illustrate how national coordination of infrastructural initiatives for the support of research into the past can generate added value, which goes beyond the sum of individual projects. Throughout the panel, the audience will be invited to join in the discussions of these projects. Interaction with the audience will bring out possible cross-connections with initiatives elsewhere in Europe, as well as potential challenges and prospects currently not addressed. In this way the panel engenders a collaborative exploration of potential transnational collaborations on the ethical, legal, organisational and technical aspects of archiving and preserving Oral History.

Topic: Interdisciplinary Approaches

Time: Wednesday, 18/June/2025: 2:00pm - 3:30pm
Session Chair: **Tanja Wissik**, Austrian Academy of Sciences

Learning from past mistakes: ACCSN 2 Project

Dilyana Boteva-Boyanova[1], Ulrike Peter[2], Lily Grozdanova[1]
[1]Sofia University "St. Kliment Ohridski", Bulgaria; [2]Berlin-Brandenburg Academy of Sciences and Humanities, Germany;
boteva@uni-sofia.bg, peter@bbaw.de, l.grozdanova@uni-sofia.bg

Learning from mistakes has always been a crucial development strategy for humanity on an individual and community level. It is a seemingly straightforward concept that is, however, severely complex.

The spread of counterfeits of ancient numismatic items leads to historical, social, and identity deformations. Recording fakes instead of originals into scientific circulation leads to endless errors with unpredictable consequences. The most effective prevention against these processes is creating a stable and proactive academic network, allowing enhanced knowledge exchange and transfer. Another central aspect is developing an IT-based digital tool that would be an active asset against the spread of objects identified as counterfeits.

The digital humanities project "Mistakes as a source of knowledge: ACCSN 2.0", supported by DARIAH-ERIC, is oriented to identifying forgeries of ancient numismatic material. This concept requires practical work and precise analysis of the counterfeits themselves. Currently, there is no digitised collection of such objects to assist specialists in further developing their expertise in production techniques and forgery markers. Hence, the scientific community's ability to study mistakes and create proactive prevention strategies is severely limited.

Building upon the achievements of the first stage of the Ancient Coins Counterfeits Scientific Network (ACCSN 1.0) development funded by the Alexander von Humboldt Foundation, the new project has several aims. The **central goal** is to digitise the most significant collection of numismatic objects identified as counterfeits belonging to the Coin cabinet of the Royal Library of Belgium. This will include on-site processing of the objects. They will be photographed and measured using a specialised QuickPX hard- and software system provided by the partner projects Corpus Nummorum (CN) and DigiThrace. After the initial processing, datasets will be created for each object and entered in Corpus Nummorum. The CN platform is currently the only digital tool actively developed, in cooperation and interoperability with ACCSN, to handle and process data for numismatic counterfeits in a scientific manner.

A further goal is to perform the second stage of the digitisation of the "Callataÿ' file of counterfeit Greek coins". It will transform the data in the info cards into DB datasets, adding them to the CN and actively connecting the stable URIs of the platform with the digital heritage objects published on the ACCSN platform.

The third goal is to develop a "standardised identification expertise card" for international application, as a severe issue is the absence of standardised expertise procedure for counterfeit identification.

This presentation proposal aims to disseminate the starting level of development, which would serve as the basis for the project. Beyond that, the presentation strives to draw more attention in the scientific community to this pressing issue and the strategies to address it.

Working across disciplines: documenting and analysing African musical instrument collections as (linked open) data

Ana Ester Tavares[1], Vera Moitinho de Almeida[1,2,3], Lucas de Campos Ramos[4,5], Jorge Castro Ribeiro[4], Rita Gaspar[6], Luís Trigo[2,7], Carlos Silva[8,7], Cláudia Oliveira[9]
[1]CITCEM – Centro de Investigação Transdisciplinar 'Cultura Espaço e Memória', Faculty of Arts and Humanities of the University of Porto; [2]CODA – Centre for Digital Culture and Innovation, Faculty of Arts and Humanities of the University of Porto; [3]INESCC - Institute for Systems and Computer Engineering at Coimbra, University of Coimbra; [4]INET-md – Instituto de Etnomusicologia – Centro de Estudos em Música e Dança, University of Aveiro; [5]CEP-EMB – Escola de Música de Brasília; [6]MHNC-UP - Natural History and Science Museum of the University of Porto; [7]CLUP - Centre of Linguistics of the University of Porto, Faculty of Arts and Humanities of the University of Porto; [8]Wikimedia Portugal; [9]InBIO- Research Network in Biodiversity and Evolutionary Biology (Associated Laboratory), CIBIO - Research Center In Biodiversity and Genetic Resources/ University of Porto; up201510202@edu.letras.up.pt, vmoitinho@letras.up.pt

In late 19th and early 20th century Portugal, art collectors and musicians sought valuable musical instruments for their antiquity, aesthetics, rarity, or provenance, reflecting a European trend for cultural heritage. However, the interest in ethnographic, historical, or "exotic" musical instruments was not limited to private art collectors but also extended to scientific expeditions and other journeys carried out in faraway territories, including European overseas colonies. This interest is also reflected through the incorporation of non-European instruments, namely, into university museums, such as the Archaeological and Ethnographic Collections at the Natural History and Science Museum of the University of Porto (MHNC-UP).

In this study, we focus on six unstudied Angolan musical instruments from the MHNC-UP with very little information available: four membranophones and two chordophones. These instruments are in good state of conservation, except for a bimembranophone and an ocihumba, a chordophone of the pluriarchs family. The first has a torn drumskin, but that doesn't hamper formal or organological analysis; the latter has some broken arches and strings, preventing a complete understanding of its functional details. Fortunately, a private collector, ethnomusicologist and musician, recently (2024) brought from Angola a functional ocihumba and direct testimonies recordings, which enable new insights into its morphology, playing techniques and instrument sound.

Hence, after a macro analysis of the physical instruments, we proceed with close-range 3D imaging (photogrammetry and 3D structured light scanning) and data post-processing; scientific illustrations are automatically generated from the high-resolution 3D models. A comparative analysis between imaging techniques is carried out, followed by specific computed measurements

and a preliminary technological and functional use-wear analysis. The woods used in the construction of these instruments are identified to understand the impact of their selection on their performance as well. Next step will consist of carrying out a spectral study of the sounds of the ocihumbas, also using the 3D digital data - i.e., to computer simulate the sounds of the ocihumba strings and body from the MHNC-UP, based on material properties and the spectral profile of the ocihumba recently brought. Following the FAIR and CARE principles, and in the spirit of Open Science , documentation is enriched with metadata, paradata, controlled vocabularies, and linked open data , into a Wikibase instance (run by CODA and hosted at FLUP) as a collaborative platform for musical instruments research.

In this presentation, we will showcase the applied workflow, while critically reflecting on the potential of interdisciplinary collaboration between various fields, namely, art history, ethnomusicology, archaeology, linguistics, botany, engineering, museology, information and documentation sciences, and the digital humanities.

ANCHISE: an interdisciplinary approach to combat illicit trafficking of cultural heritage in the digital age

Benjamin OMER[1], **Mariana Vasilache**[2], **Nikolas Kouloglou**[2], **Axel Kerep**[3], **Titien Bartette**[4], **Marine Lechenault**[5]
[1]Ecole française d'Athènes; [2]Université Lumière Lyon 2; [3]PARCS Solutions; [4]ICONEM; [5]Ecole Normale Supérieure de Police - Laboratoire de recherche; benjamin.omer@efa.gr, marianavasilake@gmail.com, nikolas.kouloglou@univ-lyon2.fr, axel.kerep@parcs-pro.com, titien.bartette@iconem.com, marine.lechenault@interieur.gouv.fr

Archaeological heritage preservation represents one of our most fundamental connections to "The Past," yet faces unprecedented threats from illicit trafficking networks that exploit technological gaps and insufficient understanding of social contexts. While numerous digital tools have emerged for heritage protection, their effectiveness remains limited due to fragmented approaches that fail to integrate social sciences with technological innovation.

The ANCHISE (Applying New solutions for Cultural Heritage protection by Innovative, Scientific, social and economic Engagement) project addresses this gap by establishing systematic collaborations between humanities researchers and technology developers, creating solutions that are both technically sophisticated and socially informed. This interdisciplinary approach recognizes that effective heritage protection requires advanced digital tools and deep understanding of human behaviors, social networks, and cultural contexts driving heritage destruction and preservation.

This paper presents two case studies demonstrating how interdisciplinary collaboration transforms understanding of heritage threats and technological responses. The first explores the intersection of sociological research and spatial monitoring technologies. Through comparative fieldwork on metal detecting practices in France and Greece, we reveal how detecting communities operate within complex social and legal frameworks varying across national contexts. These insights inform strategic deployment of appropriate monitoring technologies, with knowledge of detectorist behavior patterns helping determine whether LiDAR technology or multispectral/hyperspectral satellite imagery should be prioritized for specific sites.

The second case study explores collaboration between law enforcement agencies and artificial intelligence developers creating tools for identifying stolen cultural objects. European Law Enforcement Agencies and French Police College (ENSP) Research Laboratory members worked with ARTE-FACT image recognition system (PARCS) developers to address practical challenges of identifying potentially stolen artifacts. This collaboration revealed critical operational needs, leading to improvements in technology accuracy and law enforcement workflow integration.

Both case studies demonstrate how interdisciplinary dialogue reshapes technological design while enriching humanities research methodologies. Understanding that detectorist behavior varies across contexts demonstrates how social science research optimizes technological deployment rather than simply improving existing tools. Similarly, law enforcement perspectives revealed the importance of considering chain-of-custody requirements and legal evidence standards in AI system design.

The ANCHISE approach offers methodological innovations for digital humanities research on "The Past." It demonstrates how ethnographic and sociological methods inform archaeological site protection strategies, how law enforcement expertise enhances heritage identification technologies, and provides a framework for sustained interdisciplinary collaboration benefiting all participating fields.

This paper contributes to ongoing discussions about the role of digital technologies in preserving and studying the past by offering concrete examples of successful interdisciplinary collaboration. It addresses key topics of interest for the DARIAH community, including digital archiving and preservation, historical data analysis using AI and machine learning, spatial explorations of the past, and interdisciplinary approaches that bridge digital humanities with other fields such as archaeology, anthropology, and criminology.

By documenting these collaborative processes and their outcomes, this research provides a replicable model for future interdisciplinary initiatives addressing complex challenges in cultural heritage protection and digital humanities research.

Recreating historical figures on screen: AI resurrection, historical integrity and the ethics of representation

Jorge Franganillo
Universitat de Barcelona, Spain; franganillo@ub.edu

When bringing past events to the screen for a broad audience, archival footage often falls short of providing a complete picture. To bridge these gaps, historical documentaries and educational programs have traditionally relied on dramatic recreations and explanatory animations (cartoon-like animated sequences). However, the “digital resurrection” of deceased individuals using audio and video deepfakes is rapidly transforming the field. By training AI on archival material, filmmakers can now realistically reconstruct historical figures within audiovisual media, opening new possibilities for historical storytelling and engagement. This technological advancement, however, raises complex ethical questions regarding authenticity, representation, and the very nature of historical memory. This research explores both the transformative potential and the ethical complexities of this evolving technology, focusing on its impact on our understanding and interaction with the past.

Specifically, this study examines the implications of deepfake technology —the synthesis of realistic yet artificial human likenesses— for various stakeholders, including archivists, filmmakers, and the public. Case studies, including the AI voice cloning in *Roadrunner: A Film About Anthony Bourdain* (2021) and the digitally resurrected celebrities in *Hôtel du Temps* (2022–2024), highlight diverse approaches and ensuing ethical debates. These are contrasted with the more ethically considered recreation of comedian Pepe Rubianes' voice and likeness in *The World of Pepe Rubianes* (2024), which prioritized family consent and audience transparency, illustrating the spectrum of possibilities and the crucial role of ethical considerations.

Further, this research investigates the broader impact of generative AI on historical narratives. While acknowledging its potential to enrich understanding by creating a seemingly tangible connection to the past, it also emphasizes the risks of misrepresentation and manipulation. Responsible implementation guided by established ethical principles —such as those from the Archival Producers Alliance advocating for transparency, accuracy, and the prioritization of primary sources— is paramount, particularly concerning the accuracy and respectful representation of deceased individuals.

Finally, this study highlights the growing importance of AI literacy for both media professionals and the public. As generative AI becomes increasingly sophisticated, critical assessment and interpretation of digitally-created content is essential. This includes understanding AI's capabilities and limitations, recognizing potential biases, and engaging in informed discussions about the ethical implications. The goal is to empower individuals to distinguish between authentic historical representation and AI-generated recreations, fostering a more nuanced engagement with the past. Ultimately, this research hopes to contribute to the ongoing conversation about the responsible and ethical application of generative AI within the cultural heritage sector, ensuring its use enriches, rather than distorts, historical understanding.

Panel | MediaWiki-based tools and services in Digital Humanities workflows

Time: Thursday, 19/June/2025: 11:30am - 1:00pm
Session Chair: **David Lindemann**, UPV/EHU University of the Basque Country

MediaWiki-based tools and services in Digital Humanities workflows

David Lindemann[1], Gustavo Candela[2], Christos Varvantakis[3], Maximilian Kristen[4], Camillo Carlo Pellizzari di San Girolamo[5], Ismael Olea[6], Christof Schöch[7], Alessandro Marchetti[8], Vera Moitinho de Almeida[9], Tiago Assis[9], Alice Santiago Faria[10]
[1]UPV/EHU University of the Basque Country; [2]University of Alicante; [3]Wikimedia Deutschland; [4]Ludwig-Maximilians-Universität München; [5]Scuola Normale Superiore Pisa; [6]LaOficina Producciones Culturales; [7]Trier Center for Digital Humanities, University of Trier; [8]University of Pisa; [9]University of Porto; [10]Universidade NOVA de Lisboa; david.lindemann@ehu.eus, gcandela@ua.es, christos.varvantakis@wikimedia.de, max.kristen@kunstgeschichte.uni-muenchen.de, camillo.pellizzaridisangirolamo@sns.it, ismael.olea@laoficinacultural.org, schoech@uni-trier.de, alessandro.marchetti@unipi.it, vmoitinho@letras.up.pt, tassis@fba.up.pt, alicesantiagofaria@fcsh.unl.pt

Introduction

GLAM institutions (Galleries, Libraries, Archives and Museums) have been exploring new ways to make their content available for decades. New initiatives and practices for the publication of FAIR (Wilkinson, 2016) data have emerged, including Collections as data (Padilla et al., 2023), the CARE principles (Carroll et al., 2020) and datasheets for Cultural Heritage datasets. In this context, free open source software solutions provided by the Wikimedia Movement, particularly the collaborative linked open data platform Wikidata and the Wikibase software, have emerged as popular tools in the Digital Humanities domain (see e.g. Candela, 2024). In this regard, the DARIAH-EU Working Group DHWiki, a space for discussions and dissemination of Wikidata and Wikibase, intends to build bridges between three sectors: DH researchers, GLAM institutions, and members of the Wikimedia movement. The results of the WG are expected to be relevant input for actors like DARIAH-EU, Europeana and the ECHOES Cultural Heritage Cloud.

Wikibase is a set of extensions for MediaWiki plus other complementary services, all licensed under free libre open source licenses. It includes a data triplestore, a graphical web editor, a SPARQL endpoint and query user interface and a set of web APIs. MediaWiki is the mature software running Wikipedias in 342 languages. Wikibase is the software running the Wikidata linked open data service and community. Wikidata manages 116,337,559 different items, licensed as CC0 opendata, and 6,900,125 registered users.

As part of the work performed by DHWiki, and as a collaborative effort, a white paper is being written, covering different aspects such as FAIR data on mediaWiki-based platforms (see dedicated page and Lindemann, 2025), the use of Wikibase in DH projects, and, from a DH perspective, the identification of requirements for further development of Wikibase and related tools. More generally, this work intends to encourage GLAM institutions and DH researchers to explore the tools and services provided by Wikimedia in their FAIR publication workflows.

Format

Representing the DHWiki WG, the speakers will provide an introduction to the concepts mentioned above. This will be followed by short presentations, which loosely correspond to the chapters to be included in the planned white paper. Finally, in the public Q&A part of the panel, questions like the following can be addressed: What are requirements from a DH perspective to use an approach based on the software provided by Wikimedia, and what are requirements for further development of the platforms and related tools? How to connect the DH and Wikimedia communities? The feedback from the audience will be used for refining the white paper. This is the set of short presentations:

- DHwiki, a new DARIAH-EU working group - Welcome, introduction, and brief presentation of the new working group, its members, and its activities. **David Lindemann, Gustavo Candela**
- Wikibase feature description - Features of Wikibase, a cloud-based Linked Open Data editing and publishing solution. **David Lindemann, Ismael Olea**
- FAIR data on Wiki-platforms - Wikimedia platforms (especially Wikisource, Wikimedia Commons, Wikidata, and Wikibase) as sustainable solutions for FAIR and 5-star open data publishing. **Maximilian Kristen, David Lindemann**
- Who's using Wikibase? Examples of existing Wikibase instances in the field of GLAM and DH - Existing examples for GLAM institutions and DH research groups using Wikibase and Wikidata for different purposes, particularly highlighting MiMoText (University of Trier), a research project in computational literary studies. **Christos Varvantakis, Christof Schöch**
- Building a collaborative ecosystem for the Wikibase community - State of the art and perspective of outreach and discussion platforms and channels. **Alessandro Marchetti, Camillo Pellizzari**
- Wikidata as a hub for identifiers - The importance of Wikidata as a huge collector of identifiers, through its 9000+ external-ID properties, and how databases can benefit from interactions with it. **Camillo Pellizzari, Gustavo Candela**
- Wikibase, a solution for building transparent and collaborative data ecosystems - Wikibase for Accurate Mapping of the 'State of the Art' in Horizon Europe projects and in other EU-coordinated calls. **Alessandro Marchetti**
- MediaWiki and Wikibase as platforms for a Do-It-Together (DIT) approach to research and education in Arts - Wikibase as an Arts-Based Educational Research framework focusing on themes like environmental sustainability, social inclusion, art history and art experimentation, will be discussed based on practical examples such as eViterbo, OpusTessellatum-PT, BioImages (https://coda-hd.letras.up.pt), and Very Small GLAM. **Vera Moitinho de Almeida, Tiago Assis, Alice Santiago Faria, Ismael Olea**

Topic: Reconstructed Histories

Time: Thursday, 19/June/2025: 11:30am - 1:00pm
Session Chair: **Edward J. Gray**, CNRS

A World in Letters: Analyzing Prisoner Letters from the Early Modern Seas through Topic Modeling

Lucas Haasis
German Maritime Museum Bremerhaven, Germany; l.haasis@dsm.museum

The so-called "Prize Papers" represent a unique historical collection of documents and artefacts from ships captured by the British during the early modern period (Bevan/Cock 2018). Housed at The National Archives, UK, this extensive collection is being catalogued and digitised by the Prize Papers Project (www.prizepapers.de). A special feature of this collection is that among the records, the contents of entire mailbags have survived containing hundreds of letters. In some cases, even the corresponding jute sacks have been preserved. These collections serve as time capsules, allowing for the reconstruction of a concrete snapshot of a particular moment in history and in the lives of the writers (Haasis/Freist 2023).

This presentation will focus on letters written by prisoners who were given the opportunity to write home while on shore leave. I will show initial insights into the process of analysing letters using digital methods. What moved these writers? What was the content of the letters? What emotions are evoked? How did their social backgrounds shape their letters?

The analysis of the letters will employ a mixed-methods approach (Schneider et al 2023). In this presentation, I will concentrate on analysing the letters using the method of topic modeling, highlighting the challenges and possibilities of this technique when applied to historical letter collections (McCallum 2002, Hodel/Möbus/Serif 2022). Especially with regard to the pre-processing of the letter transcriptions, letter collections present an analytical challenge, as the letters were written by many different hands and writers, some of whom are highly literate, while others simply wrote phonetically. How can these diverse letters be standardised and thus rendered usable as a coherent sample (Bayerschmidt et al. 2025)? Furthermore, the writers come from different social backgrounds, which cannot be sufficiently captured by this method alone. The presentation will, therefore, examine how the quantitative analysis of the letters can be extended through qualitative close reading, thereby incorporating the historical context of the correspondence (Rahimi et al. 2023). Finally, it will address the overarching question of whether topic modeling produces findings that go beyond the anticipated themes one would expect to find in letters from captivity - such as reference to health, longing for the recipients or accounts of captivity experiences - in order to demonstrate the added value of the method in comparison to classical hermeneutic approaches.

The data is ready. Now what?

Pétur Húni Björnsson
The Icelandic Centre for Digital Humanities and Arts, Iceland; phb@hi.is

After three years of development version 1.0 of the Icelandic Historical Farm- and People Registry has been made accessible to the public. The registry is a research infrastructure combining data from 16 Icelandic censuses from 1703 to 1920, two mid 19th century farm surveys and two farm registries published by the Postal Service in 1885 and 1915.

The new registry has converted the source data from individual lists to networks of nodes – farms and people – making it easier to explore e.g. population and population change in individual areas, and get an overview of lifespans of Icelanders through the lens of the censuses.

A large part of the development process has revolved around cleaning and normalizing the source data, and in some instances translating it. The alignment had to be done on multiple levels as districts, counties and parishes changed in various ways throughout the period, and had to be aligned both spatially and temporally. This has called for repeated combing through the data set to seek out and rectify misalignments and other errors.

Even though the data has been deemed ready to show and allow access to it, the project is not done, and the data still has multiple errors and uncertainties, many of whom we will never be able to fix. That has proven to be another large task within the project: coming to grips with the fact that data can be considered "ready" even though it has not been fully cleaned and aligned – it's not done but it's ready.

This paper discusses the development process, the data sets involved and their state, and describes the registry's data modeling – arrived at through trials and errors – with emphasis on how the modeling lends it to external usage through API in a new project now under way, employing the farm registry as a backbone and a unified access point to various independant and often disparate databases of the Icelandic National Registry, containing information on, or connected to, farms.

Beyond the Digital: A Historical Genealogy of Virtual Reality in Western Art and Perception

Maria Eduarda Vieira Wendhausen
Faculdade de Belas-Artes da Universidade de Lisboa, Portugal; wendhausen.01@hotmail.com

Virtual Reality (VR), often perceived as a contemporary technological innovation, is deeply rooted in Western traditions of illusion, immersion, and perceptual manipulation. While recent empirical studies highlight a 230.18% surge in patent citations related to VR between 2005 and 2020, underscoring its growing significance as a critical area of interdisciplinary inquiry, this paper argues that VR is not a novel phenomenon but rather the latest iteration of a centuries-long pursuit of simulated realities. Drawing on the work of Grau (2004), Friedberg (2006), Berkman (2024) and others, this study traces the conceptual and aesthetic foundations of VR back to historical practices of illusion and representation, from the frescoes of Pompeii and Renaissance linear perspective to 17th-century peepshows, 19th-century panoramas and 20th-century cinema.

Central to this exploration is the idea that VR operates through three interconnected dimensions: the virtual, illusion, and immersion. Grau (2004) defines VR as a space of possibilities (or impossibilities) created through stimuli that deceive perception, while Friedberg (2006) emphasizes the virtual as an immaterial substitute for the material, rooted in the Platonic dialectic of image and reality. Berkman (2024) suggests that historical artefacts contributing to virtual reality often prioritize immersion through vision. This paper challenges the notion of VR as a technological breakthrough, positioning it instead as a continuation of humanity's enduring drive to simulate and reinterpret reality. By examining key historical moments—such as the proto-perspective in the Villa dei Misteri frescoes and the development of linear perspective during the Renaissance—this study reveals how ancient illusion strategies inform contemporary digital experiences.

Utilizing digital humanities methodologies, this research reinterprets historical artefacts and techniques to uncover the perceptual and aesthetic principles underlying VR. It argues that VR's immersive potential is not confined to digital media but is a fundamental aspect of human perception and representation. By situating VR within a broader historical and cultural framework, this paper not only enriches our understanding of its origins but also critiques the contemporary elevation of VR as a simulacrum in Baudrillard's (2008) sense, rather contemplates virtual and reality as twin manifestations of the same principle. Ultimately, this study contributes to a deeper understanding of VR as both a technological and cultural phenomenon, bridging the gap between historical practices of illusion and modern digital environments.

Persons in Context: Towards a European RDF vocabulary to describe person observations?

Richard Zijdeman[1,2,3], Ivo Zandhuis[1], Rick Mourits[1]
[1]IISG, Netherlands, The; [2]VU University, Netherlands, The; [3]University of Stirling, UK

This presentation communicates the applicability of Persons in Context (PiCot) vocabulary on archival records across Europe. A group of archivists and scholars across Europe have applied PiCot to various types of records in their home country. We will show how these efforts could lead to a more harmonized way of communicating person observations across countries. The outcomes are not only important to migration studies, but also to the comparability of inequality studies across multiple (European) countries. We report on the successes, failures and provide recommendations for the applicability of PiCot in a European setting. We believe PiCot could be crucial in harmonization tasks undertaken in Europeana and EOSC.

Already in ancient times, but increasingly at scale in the past 3 centuries,societies have kept records that contain information on persons. Some of these records are designed to record persons or more precisely populations, such as parish records and census records. Other records are designed for very different purposes, but still record observations of persons, such as criminal records or records on (ship's) voyages.

While many efforts have been made to utilize the information in these records, most of these efforts have been done in isolation. On the one hand there is a disciplinary gap between heritage (archives) and humanties (research) where efforts to utilize the information from these records are hardly combined. Also, within both these groups there is growth potential for coordinated efforts to provide data on person obsservations with more CARE and more FAIRly.

The Resource Description Framework (RDF) is a potential catalyst of enhanced harmonization of person observation. RDF, also referred to as Linked Data, allows for universal communication across the Web. In the Netherlands, the Center for Family History, has utilized this fact to bring a group of ontologists and the heritage and humanities togeter and create a RDF vocabulary called 'Persons in Context' (PiCot). After various iterations, this vocabulary now allows for the description of person observations for a large variety of records that exist in Dutch archives.

Topic: LLMs in Action

Time: Thursday, 19/June/2025: 2:00pm - 3:30pm
Session Chair: **Maria Ilvanidou**, Digital Curation Unit, IMSI, Athena RC

AI4LAM: A Collaborative Network for Reliable and Trustworthy Use of AI in Libraries, Archives, and Museums' Historical Collections

Ines Vodopivec
AI4LAM, National Library of Norway, Stanford University; ines.vodopivec@nb.no

AI4LAM's Activities: The AI for Libraries, Archives, and Museums (AI4LAM) community is an international, participatory network of more than 1.300 members dedicated to advancing the use of artificial intelligence within the cultural heritage sector. The community is at the forefront of developing and maintaining cutting-edge AI tools and services tailored for heritage institutions to better provide access, management and (re)use of digitised and digitally born content by supporting collaboration, innovation, and sharing of knowledge in the field of AI for institutions worldwide.

Its mission is to foster a framework for organizing, sharing and elevating the knowledge about and use of AI as well as to advocate for reliable and trustworthy AI tools and services. The community's efforts are underpinned by the principles of FAIR data, which are strongly implemented also in heritage sector by licensing frameworks, enabling fully open, interoperable and standardised metadata and rights statements which make the reuse possibilities for each item in digital collections clear.

DARIAH Collaboration: The extensive DARIAH network can play a crucial role in the development of a common dialogue environment for implementing AI technical and theoretical global developments, to be shared and adopted among LAM institutions across Europe and beyond. The AI4LAM community has already made significant strides in integrating AI technologies. This includes a range of tools and resources, such as machine learning models for metadata extraction, image recognition systems for digitized collections, and natural language processing applications for cataloguing and archival processes of historical collections. But collaboration between AI4LAM and the DARIAH communities could lead to the development of innovative solutions that address the unique challenges faced by the heritage institutions, especially when addressing issues of providing data for research of historical collections.

Use Cases of the Past: To further strengthen collaboration, this presentation will reveal the AI4LAM community's future strategic steps and showcase use cases of AI applied to historical materials, demonstrating the potential of AI in LAM institutions. By sharing experiences and insights, we can help shape the future of cultural heritage sector, research and educational stakeholders, and ensure that AI tools and resources meet the diverse needs of the upcoming decade. AI-driven solutions can enhance data management, improve access to digital collections, and streamline administrative processes. Additionally, the collaboration can foster interdisciplinary research and innovation, leading to new discoveries and advancements in various fields. The integration of AI tools into LAM digital infrastructures provides researchers with powerful tools to analyse and interpret vast datasets. This not only enhances the quality and efficiency of research but also fosters a culture of openness and collaboration across disciplines.

In conclusion, the partnership between AI4LAM and the DARIAH community represents a unique opportunity to advance the use of AI in LAM. The slogan of AI4LAM, “Individually, we are slow and isolated; collectively, we can go faster and farther,” encapsulates the community's goal: to work together and build a more innovative, secure, and collaborative future.

Archiving for the Future Past - Multimodality and AI - Challenges and Opportunities

Moa Johansson[1], Vyacheslav Tykhonov[2], Sophia Alexandersson[1], Kim Ferguson[2], James Hanlon[3], Hella Hollander[2], Jetze Touber[2], Andrea Scharnhorst[2], Nigel Osborne[1]
[1]ShareMusic & Performing Arts, 563 32 Gränna, Sweden; [2]Data Archiving and Networked Services, Royal Netherlands Academy of Arts and Science, Netherlands, The; [3]X-System Ltd, Hampshire, PO157FX England, UK; kim.ferguson@dans.knaw.nl, andrea.scharnhorst@dans.knaw.nl

This paper discusses how to enhance existing digital archival solutions with new AI-based approaches. We take as an example the creation of multimodal representations of performing arts around a newly emerging repository hosted by ShareMusic, a Swedish Knowledge Centre for Artistic Development and Inclusion. Traces of performing arts make a prime example for embracing new technological challenges when it comes to archiving in the present for the coming past. The traces usually represent complex digital objects, often combining text, image, video, 3D object representations and so on. To encapture their multimodality features as well as building multimodality (use of various senses) into retrieving them adds another layer of complexity to the digital preservation.

In this paper we present the different phases when it comes to the design of a repository fit for documentation around an inclusive performing arts with an interface providing inclusive access. Technologically, open source developments like the Dataverse project and tools to foster local implementation of mature archival solutions form the solid fundament. Leading for the design process are knowledge organisation workflows which involve human experts to create a knowledge base for arts and inclusion.

At the core of this paper, we demonstrate how innovative local AI solutions (Ghostwriter) can be used to enhance the annotation of datasets next to enhancing their accessibility via various web interface frames. In particular we zoom into the role of Monomodal Transformative AI (MTA) and Multimodal Cognitive AI (MCAI). The first (MTA) refers to a set of technologies that convert a single-source input into multiple accessible formats. For example, text can be transformed into audio or haptic representations, enabling broader accessibility for individuals with different needs. The second (MCAI) is a class of AI systems trained on multiple modalities to generate context-aware outputs by leveraging multimodal knowledge. These approaches are still in an early stage. We reflect how they can be developed further alongside the expansion of multimodal data stores, which provide the necessary corpus for effective training.

On a metalevel, this paper discusses how such innovative explorations, done in the context of EC and national funded projects (SSHOC.EU, MuseIT, SSHOC.nl) can be transported to mature repository services. Content-wise the emerging ShareMusic repository and the established Data Stations at DANS-KNAW share the fact that their collection material by nature is heterogeneous. It encompasses a spectrum from scientific documentation about humanities and arts scholarship as well as source material (of multimodal nature). A shared feature is also that 'data sets' are often produced by smaller communities either in academia and/or in society, sometimes also produced by vulnerable groups, and that the resulting traces can easily become "endangered". Adhering to the expertise function of DARIAH we exchange experiences on how to repurpose existing technological solutions and to enable division of labour via API service networks. This way costly tailored niche applications can be avoided, and the sustainability of research infrastructures for the humanities can be enhanced.

Best practices in pre- and post-ATR for historical research

Monika Renate Barget[1], Koen Hufkens[2]

[1]Maastricht University, Institute for European History Mainz; [2]BlueGreen Labs; m.barget@maastrichtuniversity.nl

Based on discussions and hands-on experiments conducted with the participants of a "Bring Your Own Data Labs" workshop hosted in Mainz in February 2025, we would like to share best practices for pre- and post-ATR (automatic text recognition) in historical research that we identified. (Barget, 2025) As ATR technologies (both for print and handwritten text) are quickly evolving due to new opportunities in Machine Learning and Artificial Intelligence, new challenges arise especially for small teams or individual researchers who may lack funding, infrastructures, expertise, or IT support to make optimal use of up-to-date tools. Moreover, existing workflows for layout and text recognition do not guarantee immediate success with historical documents in special formats or badly-preserved sources. In our workshop in Mainz, we addressed challenges from choosing the most suitable tool for one's own project to deciding what image manipulation pre-OCR and automated text cleaning post-OCR could do for researchers, depending on their research goals and work environments. (Garzón Rodríguez, 2024) The sample sources that participants brought to the workshop ranged from historical scientific records in table formats to letters, and sample documents also came in different languages. This gave us the opportunity to consider specific solutions for each case study as well as collect general recommendations. One topic that we discussed in detail was using computer vision packages for the better identification of text areas in form-like documents. Koen Hufkens (2022) shared a model workflow based on his recent work with colonial climate records from the Belgian Congo. Another important topic was the use of AI (chatbots) for image manipulation and post-OCR text correction. Here, we compared LLM tests run by Florentina Armaselu (2024) on a small selection of French-language texts with German-language tests in the DigiKAR geohumanities project. (Barget, 2023) Balancing time investment, environmental concerns and questions of research reproducibility, we found that LLMs can be successfully used to build better regular expressions or to create controlled vocabularies based on small text samples, while a direct AI-based correction of large amounts of text seemed neither sustainable nor reliable. To the surprise of some participants, we also questioned if (high-quality) OCR was always strictly necessary to answer their research questions, suggesting tools for qualitative research or image annotation software as possible alternatives. In our paper, we would like to systematically share our findings with the larger research community to invite further discussion. We believe that the topic is of considerable interest to many members of the DARIAH community as it has also been covered in the recently published DARIAH Campus training module "Automatic Text Recognition (ATR)". (Chiffoleau and Ondraszek, 2025).

Content Analysis of Historical Datasets with Large Multi-modal Models

Tianyu Yang[1], Abdallah Mohamed Abdallah Abdelnaby[2], Daniel Kurzawe[1]

[1]Niedersächsische Staats- und Universitätsbibliothek Göttingen; [2]Universität Göttingen, Germany; tianyu.yang@uni-goettingen.de, abdelnaby@sub.uni-goettingen.de

The digitizing of historical documents is a critical step in unlocking their potential for digital research. The core process of digitization involves converting scanned images of documents into a textual format that is easier to index and retrieve. Typically, a scanned document page contains not only textual content but also graphics and illustrations. Therefore, in addition to the challenges of extracting textual content from historical writings through Optical Character Recognition (OCR), the annotation of graphics and illustrations is essential. This allows these elements and their semantic content to be discoverable and analyzable through registries.

The annotation of illustrations in historical documents is a labor-intensive task traditionally handled through manual cataloging, where experts describe visual elements and assign metadata manually. More recently, specialized machine learning models have been developed to identify and classify certain types of images, such as printed illustrations or maps. However, these methods often requiring extensive training data and domain-specific fine-tuning.

To facilitate the transcription of the textual content in scanned documents, many OCR tools have been developed in recent years, e.g. PaddleOCR, EasyOCR, and so on. However, most of these tools are designed for modern documents. Historical records present greater challenges due to factors such as cursive handwriting styles, degraded text quality (e.g., faded ink or damaged paper), language changes, and complex document layouts. As a result, transcribing historical documents with general OCR tools often fails to produce accurate or fluent results.

Recently, the remarkable success of Large Language Models (LLMs), such as GPT-4, LLaMA and Vicuna has paved the way for the development of Large Multi-modal Models (LMMs), which combine pretrained visual models with LLMs to enable their visual capabilities. Trained on large scale of image-caption pair datasets, LMMs demonstrate excellent zero-shot OCR and image caption performance in the wild, which provides a valuable enhancement to the digitization workflows by automating the generation of image descriptions, complementing existing OCR-based text extraction. Their integration allows for a scalable and more efficient digitization process, bridging the gap between manual expertise and fully automated image analysis.

In this abstract, we first demonstrate the OCR and image captioning capabilities of state-of-the-art LMMs on historical document datasets. Additionally, we provide an overview of general digitization workflows and propose a feasible approach to integrate LMMs, aiming to enhance both the efficiency and discoverability of visual content in historical documents.

Topic: Let's Talk Infrastructure

Time: Thursday, 19/June/2025: 2:00pm - 3:30pm
Session Chair: **Tomasz Parkoła**, Poznan Supercomputing and Networking Center

How not to reinvent the wheel – workflows as a leverage from the past to the future

Anne Baillot[1], Megan Black[1], Massimiliano Carloni[2], Vera Maria Charvát[2], Matej Ďurčo[2], Michael Kurzmeier[1]
[1]Digital Research Infrastructure for the Arts and Humanities (DARIAH); [2]Austrian Centre for Digital Humanities and Cultural Heritage (ACDH-CH), Austrian Academy of Sciences (OEAW); anne.baillot@dariah.eu, megan.black@dariah.eu, massimiliano.carloni@oeaw.ac.at, veramaria.charvat@oeaw.ac.at, matej.durco@oeaw.ac.at, michael.kurzmeier@dariah.eu

In recent years, a significant range of digital resources and methodologies have been developed in the Arts & Humanities. By reusing these resources, researchers can minimize redundancy and foster greater collaboration. However, challenges arise when methods are difficult to adapt or reflect biased perspectives. A key solution to these challenges lies in the thorough documentation of research choices, which ensures reproducibility and allows future generations to build on prior work. This concept is especially vital in workflows, which serve as a critical means of recording and reproducing the 'past of research'. The workflow descriptions featured in the SSH Open Marketplace (SSHOMP) play a central role in capturing this essential documentation, forming the primary focus of this paper.

We intend to examine the impact of various workflow paradigms on research reproducibility. Workflow typologies extend across a spectrum, ranging from text-based descriptive methodologies as employed in the SSHOMP to fully executable code-based frameworks, with intermediary hybrid forms – such as those employed in the Journal of Digital History – integrating features of both approaches. Although workflows serve as invaluable instruments for structuring research methodologies, their efficacy in ensuring research documentation and methodological reproducibility is contingent upon both their type and the contextual environment in which they operate.

Descriptive, text-based workflows as described by Barbot et al. afford greater structural and conceptual flexibility, functioning as high-level expositions rather than direct computational scripts. They facilitate the articulation of abstract methodological frameworks, offering enhanced accessibility to both authors and readers, while they minimize the technical overhead associated with their maintenance, thus also ensuring their long-term viability. However, more rigorous editorial oversight to maintain their scholarly integrity and usability is needed.

Conversely, code-based workflows provide a structured and automated approach to research reproducibility by leveraging computational scripts to execute analytical procedures. They enable precision, scalability, and automation, mitigating the risks of human error inherent in manual documentation. Moreover, they can facilitate seamless integration with version control systems, enhancing transparency and collaboration across research teams. However, code-based workflows require domain-specific technical expertise and are susceptible to software dependencies, which may impede their long-term accessibility and interoperability (Clavert et al.).

As workflows increase in executability, they become more enmeshed with specific software ecosystems, heightening their risk of deprecation as platforms evolve or become obsolete. This introduces a paradox: while greater executability enhances methodological rigor and repeatability, it may prevent the reconstruction of past research due to shifting technological landscapes.

This paper will systematically categorize different workflow types, with a particular emphasis on those utilized within DARIAH, created in the context of the ATRIUM project, to assess their suitability for documenting research processes. By analyzing how workflows function within scholarly infrastructures, it will offer insights into best practices for ensuring their longevity and accessibility. Furthermore, the study will provide recommendations for designing workflows that balance methodological rigor with sustainability, thereby making them more effective as tools for preserving and interpreting the past of research.

Exploring the past with the AVOBMAT (Analysis and Visualization of Bibliographic Metadata and Text) multilingual research tool

Róbert Péter, Zsolt Szántó, Zoltán Biacsi, Gábor Berend, Vilmos Bilicki
University of Szeged, Hungary; robert.peter@ieas-szeged.hu

he objective of this paper is to introduce the AVOBMAT (Analysis and Visualization of Bibliographic Metadata and Texts) multilingual research tool, which enables researchers to critically analyse bibliographic data and texts at scale with the help of data-driven methods supported by Natural Language Processing techniques. This exploratory tool offers a range of dynamic text and data mining tasks and provides interactive parameter tuning and control from the preprocessing to the analytical stages. The analysis and visualization tools both facilitate close and distant reading of texts and bibliographic data.

Compared to other similar tools, the unique features of the AVOBMAT toolkit are: (i) the use of transformer language models on a scalable, cloud-based infrastructure that allows researchers to preprocess and analyse texts and metadata at scale; (ii) it combines metadata and textual analysis, enabling users to ask complex research questions, in one integrated, interactive and user-friendly web application; (iii) it analyses and enriches texts and metadata in 16 languages; (iv) at the preprocessing phase, texts can be cleaned and each analytical tool can be individually configured using a total of 19 parameters; (v) users can test, validate, and save the configuration settings; (vi) private databases can be made public.

The user can search and filter the metadata and texts in faceted, advanced and command-line modes and perform all the subsequent analyses on the filtered dataset. AVOBMAT offers the following analytical functions: (i) metadata analysis (line, area, bar, pie network analyses); (ii) lexical diversity analyses; (iii) n-gram viewer; (iv) topic modelling; (v) frequency analyses (keyword context, significant text analysis; (vi) named entity recognition, disambiguation and linking (Wikidata, ISNI, VIAF), (vii) part-of-speech tagging; (viii) keyword-in-context.

The reproducibility and transparency of the experiments and results using the tool are enhanced by the ability to import and export the parameter settings as templates or JSON files. Users can create templates for the preprocessing and analytical functions on the graphical interface. The tabular statistical data and visualizations of the performed analyses can be exported in PNG or various CSV formats.
AVOBMAT helps users explore large historical and literary collections, uncover novel insights into historical events and trends, and unveil overlooked connections, themes and patterns. As sample databases, we have preprocessed the DraCor dramas (3793 in number) and ELTeC novels (1505) in 13 languages. We enriched the metadata of these collections and corrected inaccuracies. For example, as for the DraCor collection, we added, among others, the list of characters with their gender annotations, the authors' gender and their age at the time of writing.

The beta version of AVOBMAT will be available via the infrastructure of the Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG).

Towards interdisciplinary approaches to digital cultural heritage: GLAM Labs and data spaces

Alba Irollo[1], Gustavo Candela[2], Mahendra Mahey[3], Katrine Hofmann[4], Olga Holownia[5], Nele Gabriëls[6], Steven Claeyssens[7]

[1]Europeana Foundation, Netherlands; [2]University of Alicante, Spain; [3]Tallinn University, Estonia; [4]Royal Danish Library, Denmark; [5]International Internet Preservation Consortium, USA; [6]KU Leuven Libraries, Belgium; [7]National Library of the Netherlands, Netherlands; alba.irollo@europeana.eu, gcandela@ua.es, mahendra.mahey@tlu.ee, khg@kb.dk

At the intersection between the cultural heritage and research sectors, the increasing availability of digital cultural heritage data has supported new ways of producing knowledge, research, publishing results, and teaching in academic settings. The efforts put into the digitisation of their collections and capturing born digital archives, have led GLAMs (Galleries, Libraries, Archives and Museums) and other organisations such as universities to focus on many factors. These include the wider accessibility of the data and its reuse-related potential refining approaches to digital curation so that digital cultural heritage can support computational analysis, especially through Digital Humanities approaches. Emergent initiatives, such as Collections as Data, FAIR and the CARE principles for Indigenous data governance, have emphasized the need for best practice when making data available based on sets of principles (Padilla, 2023; Carroll, 2020). Moving from these principles, the International GLAM Labs Community has grown as a collaborative and interdisciplinary initiative aimed at promoting the publication, computational access and responsible use of data (Mahey, 2019; Candela, 2023; Candela, 2023b). It includes more than 90 institutions covering a wide diversity of domains from GLAMs and wider. In this context, collaborations with researchers and university support staff from different fields have become crucial factors in supporting the study of the past with a careful eye on the needs of the present and future.

- The European data space for cultural heritage fosters collaborations between the cultural heritage sector and academia, building on the experience of the Europeana Initiative and its long-standing partnerships with research infrastructures like DARIAH. These partnerships have led to joint efforts, especially in digital humanities. A new trend deserves attention in this context, and it is the increasing number of university courses focusing on digital cultural heritage across Europe. Acting as an observatory, the data space aims to foster exchanges around this topic, enhancing cooperation between the academia and cultural heritage institutions in the educational paths of the next generation of (digital) curators.

- GLAM institutions are exploring new ways to make their content available suitable for computational access. This section will introduce examples of publication and reuse of digital collections published by relevant institutions such as the Royal Danish Library and KU Leuven. It will include a description of challenges and opportunities illustrating their approach.

-The publication of digital collections suitable for computational use can be a difficult task for the institutions. Best practices and guidelines to publish digital collections suitable for computational use promote its adoption (Candela, 2023a). This section will introduce a selection of relevant and innovative initiatives that can be useful in this context. It will also provide examples of use based on a wide diversity of content such as GLAM datasets and Web Archives.

- Will highlight examples of use and lessons learned from his research with users of digital cultural heritage at the British Library and those providing services for the reuse of digital cultural heritage as data around the world in the GLAM Labs community.

Topic: Transforming Digital Methods

Time: Friday, 20/June/2025: 9:30am - 11:00am
Session Chair: **Nanette Rissler-Pipka**, Max Weber Foundation

The Digital Transformation of Maya Hieroglyphic Research

Christian Prager
University of Bonn, Germany; cprager@uni-bonn.de

The classification and analysis of hieroglyphic writing systems present methodological challenges within Digital Humanities. Focusing on the Classic Maya script, this paper examines the limitations of traditional iconographic approaches in addressing the graphic and semantic complexity of Maya hieroglyphs. Established classification systems, such as J. Eric S. Thompson's catalog, remain fundamental references but exhibit methodological constraints due to overlaps between iconographic and semantic criteria and the static nature of printed catalogs, which hinder updates and integration of new discoveries.

To address these challenges, the "Text Database and Dictionary of Classic Mayan" project refines and extends Thompson's classification system through digital methodologies. A key innovation is the systematic digital documentation and encoding of palaeographic variants, particularly anthropomorphic and zoomorphic glyphs, historically underrepresented in classification efforts. By employing a numeric coding system independent of iconographic descriptions, this initiative provides a flexible framework, mitigating limitations of static classifications and enabling more precise analyses of Maya hieroglyphs' formal, semantic, and functional dimensions.

Another central aspect is the implementation of controlled vocabularies for consistent iconographic descriptions within digital research environments. This system supports structured analyses based on both external morphological traits and internal semantic properties. Additionally, the digital catalog framework facilitates the integration of newly identified glyphs, while digital concordances enable transparent comparisons with earlier classification systems. Researchers can systematically evaluate historical cataloging efforts in relation to contemporary findings, refining methodological perspectives.

A crucial feature of the project is its integration of TEI (Text Encoding Initiative) and XML standards for encoding textual data. The platform serves as an interface between RDF-based data structures and TEI-encoded textual content, enabling seamless retrieval and visualization of hieroglyphic information. Texts stored in TEI format are dynamically incorporated into the research portal, ensuring structured textual data can be efficiently linked with broader semantic web technologies. This interoperability facilitates data exchange between digital resources, creating interconnected research workflows.

Through case studies of recent discoveries, this presentation demonstrates how digital methodologies address classification constraints while maintaining continuity with established frameworks. Examples illustrate the benefits of a digital approach in classifying and analyzing variant glyphs, offering deeper insights into their linguistic, cultural, and functional contexts. Additionally, the study highlights digital interoperability's role in fostering collaborative research and enhancing accessibility to Maya hieroglyphic studies.

Beyond Maya epigraphy, these methodological advances provide a model for the digital analysis of other complex historical writing systems facing similar challenges. By situating the discussion within broader Digital Humanities debates, this paper encourages interdisciplinary collaboration and underscores the necessity of robust digital infrastructures for ongoing research. Ultimately, this study demonstrates how digital innovation enhances the analysis and dissemination of historical scripts, fostering new perspectives in epigraphy and cultural heritage studies.

Valorizing Past Art Historical Research with LLMs for European Cultural Heritage: A Case Study of the Corpus Rubenianum

Arnoud Wils
Maastricht University, Netherlands, The; arnoud.wils@maastrichtuniversity.nl

The vast body of past art historical research, including encyclopaedias, monographs and art catalogues, represents a wealth of meticulous bibliographic and scholarly research. However, a significant portion of this valuable material exists only as scanned images or unstructured OCRed PDF documents, posing a challenge to contemporary researchers seeking to fully exploit these findings. This limitation hinders the seamless integration of established knowledge with newer discoveries and state-of-the-art digital analysis methods.

This paper explores an approach to bridge this gap by **demonstrating the potential of Large Language Models (LLMs) to extract structured information from these historical resources**.

To illustrate this potential, our research focuses on the *Corpus Rubenianum*, an impressive catalogue of over 40 books meticulously documenting the work of Peter Paul Rubens. This monumental work, based on the lifelong research of Ludwig Burchard, is universally recognised as the definitive resource on Rubens. Each volume is written by a leading scholar and aims to embody all that is currently known about the artist's oeuvre, with over 2,500 compositions and 10,000 works of art listed, based on Burchard's extensive documentation.

Through the use of LLMs, we aim to extract key structured data from a selection of the volumes of the corpus available in PDF only, including but not limited to

- structured bibliographic references
- a structured list of provenances for each artwork
- a structured list of Greek mythological figures depicted in each painting (iconography).

The second part of this paper will demonstrate how this extracted structured data could be effectively valorised through various digital methods, including but not limited to:

- enriching current bibliographic information with structured extracted bibliographic references from PDF sources.
- creating an index of characters depicted on the paintings with pointers to the respective artworks.
- visualising structured provenance information for specific artworks in an interactive timeline or network graph of individuals or institutions that owned an artwork during a particular period.
- illustrating the potential of feeding the structured iconographic descriptions into a vision transformer model to develop an augmented reality layer that displays iconographic information alongside the artwork, allowing a wider audience to 'read' the artwork.

A prototype demonstrating these LLM extraction capabilities was recently presented and awarded second place at the ai4culture hackathon (in partnership with europeana.eu - February 2025).

By demonstrating the extraction of structured information from a significant historical resource such as the *Corpus Rubenianum*, and its subsequent digital manipulation and visualisation, this paper will highlight the transformative potential of these technologies to unlock data from valuable past research that is only available in unstructured PDF-like formats. This approach can facilitate more comprehensive research and enhance public engagement through digital interfaces.

Topic: Databases, from Past to Future

Time: Friday, 20/June/2025: 9:30am - 11:00am
Session Chair: **Amelia McConville**, DARIAH-EU

Teaching Literary History through Computational Analysis

Marie-Christine Boucher, J. Berenike Herrmann
Universität Bielefeld, Germany; marie-christine.boucher@uni-bielefeld.de, berenike.herrmann@uni-bielefeld.de

Understanding literary history is a fundamental component of the literary studies curriculum (Summit 2010). Traditionally, students have been introduced to the discipline from a historical perspective, often following a linear temporal progression that outlines literary periods, or simply using centuries and national or linguistic boundaries as markers. In many academic settings, such as in Germany, students of literary studies often also pursue teacher training. Consequently, they learn about literary history as future mediators of a constructed (national) literary tradition. However, as many critics have pointed out in recent decades, doing literary history poses certain challenges: Which texts are read and which are forgotten? Who becomes part of the canon, and how does this intersect with power relations? What kind of literary history are we teaching if we only address a very small part of the literary archive?

Outside of the increasingly popular but not yet widespread digital humanities programs, literary history is generally taught through the lens of hermeneutics, or, less frequently, social history. Complementing these, digital humanities and computational methods can offer promising solutions to questions about large-scale phenomena over long periods of time. By working quantitatively with large data sets—some call them *capta* to emphasize their constructed nature (Drucker 2011)—computational literary studies strive to "explain, or to provide, general laws of literature, and even of history and culture" (Bode 2023, 547).

While scalability in data analysis offers broader perspectives, it does not eliminate the need for narrativization. Any literary history inherently synthesizes information and reduces complexity. Expanding the data set accentuates potential issues related to data selection (Herrmann and Lauer 2018). So the process of building a corpus doesn't eliminate the challenges of hermeneutic or social-historical literary history. Rather, the practices of quantitative research tend to expose problems around canonicity, periodization, and intersectionality, by making the selection process transparent. The same holds true for diagrammatic visualizations, which play a crucial role in quantitative analysis. They explain relationships and developments within literary history by providing visual frameworks for concepts of time and change, causality, and continuity (Börner et al. 2016). Therefore, a key advantage of computational literary studies is its ability to identify and address selection processes and blind spots (Herrmann et al. 2025), while critically engaging with issues like availability, selection, bias, and canonicity in the narrativization of literary history (Underwood 2019).

We discuss these questions using an interactive Open Educational Resource (OER) in data literary studies as a case study. This OER introduces concepts such as modeling, operationalization, corpus building, and various measures of quantitative analysis to students and teachers of literature who are not yet familiar with digital humanities methods. Through this example, we highlight how fostering data literacy in students equips them with the skills to critically engage with narrative constructions of literary history.

SHEWROTE database launch: Past lessons and future challenges redeveloping a heritage database

Alicia Montoya[1], Pia van de Schaft[1], Viola Parente-Čapková[2], Marie-Louise Coolahan[3], Nina Geerdink[4], Katja Mihurko[5], Nicole Pohl[6], Emmanuelle Radar[7], Amelia Sanz[8], Marie Nedregotten Sørbø[9], Jasmine Westerlund[2]
[1]Radboud University (The Netherlands); [2]University of Turku; [3]NUI Galway; [4]University of Utrecht (The Netherlands); [5]University of Nova Gorica (Slovenia); [6]Bodleian Libraries, Oxford University; [7]Leiden University; [8]Complutense University of Madrid; [9]Volda University College (Norway); alicia.montoya@ru.nl, pia.vandeschaft@ru.nl

How does one revive an older database once the software it was built with is no longer supported, and after project funding run out? How does one honour and preserve the work of past scholars while making the database fit for future generations? In this presentation, the formal launch of the SHEWROTE database, we discuss our experience redeveloping a heritage database originally created in 2001, as Women Writers (www.databasewomenwriters.nl), revamped in 2014 as NEWW: New Approaches to European Women's Writing (https://womenwriters.rich.ru.nl/womenwriters/vre/), and finally redeveloped as SHEWROTE (Studying Historical Early Women's Reception: Oeuvres, Texts, Engagements) (https://shewrote.rich.ru.nl/), from 2023 onward.

The evolution of the database over decades, including several major technical overhauls, meant that research questions, scope, and (implicit) ontologies also changed significantly between 2001 and today. The data in the first versions of the database was uneven in geographic and temporal scope, and displayed gaps resulting from the different research questions historians had sought to address with it. The single biggest challenge, however, was the lack of a formal data model, and ontologies that had been developed implicitly and heuristically by past generations of researchers. Our first major decision in reviving the database, therefore, was to rethink and restructure the data completely, from the ground up, moving from a non-SQL structure (a SORL platform) to a SQL one (a Django solution).

The old databases provided an exceptionally rich dataset detailing the many forms that the reception of women writers and their works had assumed from Antiquity to circa 1945. This data had been structured, over the years, not from decisions made in a 'top-down' data model, but bottom-up, as researchers heuristically described the material they were dealing with when collecting data for the first iterations of the database. But between 2001 and 2023, two new digital projects addressing women's authorship, both offshoots of the Women Writers project, the RECIRC (https://recirc.universityofgalway.ie/) and MEDIATE databases (https://mediate-database.cls.ru.nl/about/), revealed important new facets of our type of data. This included the need to create a separate, new field distinct from Reception: Circulation, or the historical movement of physical copies of works by women writers, as attested for example by private and institutional libraries, across time and space.

In this presentation, we discuss how we modelled, implemented and populated the new Circulation field, using both data already present in the old database, and importing new datasets from other databases. With modern geo-referencing tools and analytic

computing power now at our disposal, our ability to visualize the physical circulation of books – for example, through time-lapse maps illustrating the geographic movement or spread of specific works – has expanded exponentially. At the same time, creating such visualizations underlines other significant data gaps, including North-South data inequities and the lack of historical gazetteers documenting historical place names and the changing administrative entities with which they were associated throughout the long period covered by the SHEWROTE database. Although the public version of the database has now been launched, therefore, future challenges remain ahead that our project will continue to tackle.

DigitalSEE: Mapping History and Cultural Identity

Kristiyan Sergeev Simeonov, Maria Baramova
Sofia University "St. Kliment Ohridski", Bulgaria; sergeevs@uni-sofia.bg, baramova@uni-sofia.bg

DigitalSEE is a comprehensive multimodal project dedicated to extracting and structuring historical data from diverse sources such as Ottoman travelogues, woodcuts, and archival materials. Its primary data sources include the renowned collections of Felix Kanitz, Karel Škorpil, and Konstantin Ireček. The project employs a non-proprietary, well-documented XML format with capabilities for exporting information in TEI XML format, while future developments aim to integrate the CIDOC-CRM standard for enhanced interoperability. By meticulously tracing sources and analyzing both textual and visual content, DigitalSEE focuses on the perception and reception of artifacts and monuments across time.

The objective of DigitalSEE is to conduct a diachronic study of cultural heritage and identity in the Balkans during the 18th and 19th centuries. The project places special emphasis on historical frameworks derived from Antiquity, the Middle Ages, and the Ottoman Period. Such an approach facilitates a nuanced understanding of the evolution of cultural landscapes and national identities, especially in light of the Eastern Question and the nation-building processes in Southeastern Europe during the 19th century. The integration of European travel writings, diplomatic records, and cartographic sources—particularly along the Via Diagonalis and the Danube—further enriches the study by providing context and detail about the region's historical transformation.

At the technical level, DigitalSEE is powered by a Python Flask web application that serves as the backbone for data processing, input, storage, and retrieval. This application features a form-based input system that accepts text, images, and geographic coordinates, linking data to geospatial databases such as GeoNames and Pleiades. The platform supports data storage in XML and JSON formats, while offering map visualizations that enable users to explore the historical data interactively.

The workflow underpins the project, with custom-built XML and JSON master files ensuring efficient data management. The system is designed to convert and preserve historical information in a standardized TEI XML format. Furthermore, the data structure incorporates modified metadata encoding to capture heterogeneous source properties, including detailed dating, provenance, and geospatial information. Elements such as detailed descriptions of materials, dimensions, and architectural features are used to reconstruct a broader archaeological and historical context.

Future plans for DigitalSEE include the integration of a specialized image recognition model based on a curated image dataset and a pre-trained model. Advanced topic modeling algorithms will be employed to identify clusters of similar words and facilitate research-assisted interpretation of the texts. In addition, a multimodal pipeline is under development to process sources in 18th-century Latin, German, and Old French. This pipeline leverages transcription tools like Transkribus and Kraken, along with natural language processing libraries such as spaCy's LatinCy and the Classical Language Toolkit. The resulting data, enriched with metadata and linked to external resources, will further support spatial analysis and visualization in GIS applications, ensuring a comprehensive exploration of the Balkans' historical heritage.

Unlocking the Past: The Biblissima Portal, a Gateway to Ancient Written Heritage in the Digital Age

Emmanuelle Morlock[1], Anne-Marie Turcan-Verkerk[2,3], Régis Robineau[2,3], Eduard Frunzeanu[2,3], Kevin Bois[2,3]
[1]CNRS, France; [2]EPHE-PSL, France; [3]Campus Condorcet; anne-marie.turcan-verkerk@ephe.psl.eu, regis.robineau@biblissima-condorcet.fr

The Biblissima portal is a discovery tool for specialized digital resources in the field of ancient written heritage, ranging from the earliest Mesopotamian clay tablets to the first printed books.

It is the centrepiece of Biblissima's infrastructure, a project for the creation of a digital research infrastructure that has been funded by the French Government since 2012. The portal was launched in 2017 and has since continuously ensured the interoperability of diverse and complementary data sources.

The portal thus aggregates digital data on early manuscripts, incunabula and printed works, and provides access to digitised documents, catalogues, scientific databases, illuminations, an iconographic thesaurus and soon electronic editions.

The Biblissima portal stands out as an innovative tool for several reasons:

- It provides a unified access point to heterogeneous digital resources, serving both the general public and the most specialised researchers in the field.
- It enables the post hoc aggregation and interoperability of disparate datasets, building a reference tool that no single source provider could produce independently.
- It utilises the International Image Interoperability Framework (IIIF) standard for disseminating images online, allowing direct access to primary document images, along with advanced features where available (such as full-text search, content indexes, and annotations).
- It offers innovative, user-friendly search and navigation interfaces, including an iconographic exploration tool, providing a structured and efficient starting point for researcher's investigation while encouraging critical reading of the search results without masking the shortcomings of the source databases or their contradictions.

The portal facilitates direct access to resources that are otherwise difficult for the general public to reach. It also establishes itself as a reference tool for research in the field of ancient written cultures, whose authority is being built up over time, in conjunction with the other research communities that develop components of the broader infrastructure (such as tools for analysing and processing primary data, such as the eScriptorium HTR platform). It thus plays a stabilising and consolidating role for each of its data sources, while providing undeniable added value in terms of international visibility.

To illustrate how the Biblissima portal has become an essential tool for research into ancient written cultures, the talk will be divided into three parts. The first part will demonstrate a typical research journey within the discovery tool, showing how search results are arranged to facilitate connections and entry points for launching a critical investigation. This quick overview will then provide a better understanding of the data interoperability method, its issues and its challenges. The final section will focus on the roadmap currently being drawn up to enhance the usefulness and authority of the portal. This includes improvements to search and navigation interfaces and the development of a coordinated “data governance model” involving source database contributors who wish to participate in the portal’s evolution. Through these efforts, Biblissima aims to strengthen its role as a research infrastructure coordinating the development of a key reference tool in the field of ancient written cultures, fostering cross-research opportunities, knowledge consolidation, and new discoveries.

Demonstrations

Time: Thursday, 19/June/2025: 9:30am - 11:00am
Session Chair: **Kim Ferguson**, DANS

Teaching the Past with Future Tools: Digital Humanities in Historical Education

Mojca Šorn[1], Neja Blaj Hribar[1], Ana Cvek[1], Ida Leonida Gnidovec[1], Vojko Gorjanc[1,3], Nataša Henig Miščič[1], Tjaša Konovšek[1], Tamara Logar[1], Katja Meden[1,2], Mihael Ojsteršek[1], Kristina Pahor de Maiti Tekavčič[1,3], Sergej Škofljanec[1], Robert Vurušič[1]
[1]Institute of Contemporary History, Slovenia; [2]Institut Jožef Stefan, Ljubljana, Slovenia; [3]Faculty of Arts, University of Ljubljana, Slovenia; ana.cvek@inz.si, ida.gnidovec@inz.si, katja.meden@inz.si

The Research Infrastructure of the Institute of Contemporary History (coordinating entity of DARIAH-SI) has long offered practical training in areas such as DH and library and information science.

In 2024, a new initiative was launched in cooperation with the Department of History, Faculty of Arts, University of Ljubljana, which aimed at providing a structured training programme for students. The result was a comprehensive course focussing on 19th century research, combining in-depth lectures with practical training to enhance students' historical research skills, in particular with digital methods

The course was divided into four phases:

1. Introduction to Research Infrastructure: students learnt about the research landscape in Slovenia, focusing on the Institute's role, covering the entire data cycle, metadata processing and publication on the SIstory portal.
2. Foundations of digital humanities: the next phase provided an overview of the field, covering its key definitions, methods and various outputs such as digital editions, databases, tools and corpora. Students were introduced to the digitisation pipeline, from raw content to structured corpora, along with basic XML and TEI encoding principles.
3. Hands-on training: practical sessions reinforced theoretical knowledge. Students worked with individual TEI XML files containing errors in the transcriptions, correcting them directly within the TEI framework. Additionally, they were introduced to noSketchEngine, where they followed search processes and explored linguistic data.
4. Individual research projects: in the final phase, students applied their newly acquired skills to individual seminars.

For their research seminar, they could choose from the following five different modules:

- Carniola Regional Assembly (corpus Kranjska 1.0): this module introduced students to the stenographic records of the Carniola Regional Assembly and allowed them to analyse the debates by tracking themes like nationalism using key terms. A corpus-based analysis was to be contextualised with historical positioning, literature and newspaper sources.
- Stenographic records of the First Yugoslavia (corpus yu1Parl 1.0 (1919-1939)): this module enabled research on specific topics and keywords within the parliamentary debates of the Assembly of the First Yugoslavia (1919-1939). Although this period extends beyond the 19th century—the course's focus—it allowed for comparative analysis with other parliamentary corpora.
- Population censuses: this module allowed students to familiarise themselves with censuses and the SIstory transcription tool by transcribing and analysing some of the data. The latter focuses on the methodology of processing census data and its applicability in historical studies.
- 19th and 20th century personalities: the module enabled students to study the 19th- and 20th- century personalities associated with the Carniolan or Yugoslav assemblies (up to the 1920s) by analysing their work using archival sources, newspaper archives and relevant corpora for additional context.
- Content addition to the History of Slovenia – SIstory portal: students processed the publications assigned to them through the entire data workflow, from acquisition to publication on the SIstory portal. They also analysed the publication in a broader historical context, using newspaper archives and relevant literature.

Each student worked with two mentors proficient in historical research methodology and digital techniques, respectively. This ensured a well-rounded approach to theory and methodology, preparing students to integrate digital methods into historical research.

How to annotate those thousands of entities? Approaches to (semi-)automatic entity linking for scholarly editions.

Felix Helfer[1], Thomas Eckart[1], Uwe Kretschmer[1], Johannes Korngiebel[2], Martin Prell[2], Margrit Glaser[2]
[1]Saxon Academy of Sciences and Humanities in Leipzig; [2]Klassik Stiftung Weimar; helfer@saw-leipzig.de

Entity Linking is the resolution of entity mentions to appropriate entries in a knowledge base. It represents a valuable, albeit laborious enrichment for text-based data. Our poster presents work-in-progress experiments to annotate texts of the project PROPYLÄEN, which is a research project of the Saxon Academy of Sciences and Humanities in Leipzig in cooperation with the Klassik Stiftung Weimar/Goethe and Schiller Archive and the Academy of Sciences and Literature |Mainz, with linked entities and the practical application of this enriched data in an entity-based, federated search engine.

The poster introduces the PROPYLÄEN project, which merges biographical data of Johann Wolfgang von Goethe (letters, diary entries and other testimonials) from four different sub-projects, enriches them and makes them available digitally in a fifth sub-project – most prominently on the project's research platform (https://goethe-biographica.de/). In the digital texts, mentions of people and places should be linked to entries in the research database so:fie, preferably with automated or semi-automated processing. Initially, this will be tested for a subset of the data for which a register exists documenting all entities *occurring* in the text – but not directly annotating them.

The poster discusses experiments by the SAW Leipzig for the linking process for this subset of data. Entity linking can be divided into three sub tasks: First, the detection of entity mentions in the text, which is often achieved via named entity recognition. Second, a candidate search, meaning the preselection of a subset of suitable candidate entries in the respective knowledge base to shrink the search space. In this case however, the register already predefines the candidate sets. Third, the candidate disambiguation, which for every entity mention ranks the candidate entries to find the most likely link. For this disambiguation, several research questions are explored in experiments or previewed: Does a string-based matching (using edit distance or a similar metric) already give usable results? Can an embedding-based approach, drawing from additional information in the knowledge base, improve on it in a meaningful way? Can these embeddings be improved with additional information, e.g. from external knowledge bases like the Integrated Authority File / Gemeinsame Normdatei (GND) or Wikidata?

These experiments are relevant beyond the context of the PROPYLÄEN project: Especially in retrodigitized resources, there often are registers available for the relevant entities in a resource – but without the explicit annotations in the text itself. Therefore, an application for automatically annotating these resources – even with a possible human-in-the-loop quality assurance – could help enrich them in a more resource-efficient manner.

Finally, a practical application of EL-annotated data is introduced: the integration of the annotated resource into the *EntityFCS* – a federated, entity-based content search platform of the German research infrastructure consortium Text+. This allows users to query relevant text passages with entity IDs of the appropriate knowledge base – showcasing how these annotations can increase, among other things, the explorability and findability of a resource in the context of a research infrastructure.

Interdisciplinary Approaches in the Dariah.hub Poland e-infrastructure

Krzysztof Abramowski[1], Aleksandra Nowak[1], Marcin Heliński[1], Bartosz Szymendera[1], Tomasz Umerle[2], Tomasz Parkoła[1]
[1]Poznan Supercomputing and Networking Center, Poland; [2]The Institute of Literary Research of the Polish Academy of Sciences, Poland; anowak@man.poznan.pl

Dariah.hub (2024-2025) aims to deliver a new, integrative platform for DARIAH-PL: the Interdisciplinary Research Platform. Thanks to Dariah.hub, the Polish consortium is able to leverage the infrastructure of complementary and distributed laboratories set up in a previous Dariah.lab project (2021-2023) into a new service: a platform to integrate state-of-the-art digital methods from various disciplines of digital humanities. By uniting diverse disciplines, the platform expands research perspectives, encourages knowledge sharing, and fosters synergistic collaborations.

The platform is designed to facilitate interdisciplinary research by leveraging multidimensional data models, allowing for integration across spatial, temporal, and behavioral dimensions. This enables a more holistic approach to source material analysis, fostering connections between previously unlinked datasets and methodologies. Through advanced tools like Optical Character Recognition (OCR), Handwritten Text Recognition (HTR), and Named Entity Recognition/Linking (NER/NEL) – as well as a broader suite of research tools integrated at multiple levels, the platform empowers researchers to conduct cross-domain analyses with increased accuracy and speed. A core feature of this infrastructure is its modular architecture, which streamlines comprehensive data management of diverse files, including texts, images, and multimedia repositories relevant to multiple disciplines. The platform is integrated with multiple tools from Dariah.lab toolkit, enabling cross-referencing of various sources, ensuring interoperability across disciplines. These tools are complemented by collaborative workspaces, where secure, shared editing, and transparent version control enable seamless multi-author engagement. Moreover, automated workflows and data pipelines facilitate interoperability between different tools, ensuring that researchers can connect diverse datasets without disciplinary or institutional limitations. Open licensing frameworks support data sharing while reinforcing interoperability standards crucial for large-scale, cross-institutional investigations. This ensures that research outputs remain accessible to a wide range of stakeholders, including academia, cultural institutions, and commercial entities interested in digital humanities applications.

At the heart of this endeavor is a dynamic feedback loop between data providers (e.g., archivists, cultural institutions, field researchers) and data consumers (e.g., historians, sociologists, anthropologists, archaeologists). This iterative exchange drives continuous enhancement of curated datasets, simultaneously expanding the underlying knowledge graph. Drawing upon project partners long-standing tradition in high-performance computing, the platform's computational backbone manages resource-intensive tasks – from large corpus analyses to 3D reconstructions of archaeological sites – opening new lines of inquiry for scholars across an ever-wider range of fields.

By integrating advanced digital tools with established scholarly practices, the Interdisciplinary Research Platform paves the way for explorations that transcend conventional academic boundaries. Archaeologists and anthropologists alike can combine textual and spatial data to delve into cultural patterns, while sociologists utilize networked archives to contextualize both historical and contemporary social phenomena. Ultimately, this collaborative environment not only broadens the scope of digital scholarship, but also showcases how genuine interdisciplinary synergy can deepen and enrich humanities research at large.

Enhancing the Digital Humanities Research in R: Accessing the Finnish Cultural Heritage Data through R Packages finna and finto

Akewak Jeba, Julia Matveeva, Leo Lahti
Turku University, Finland; akjeba@utu.fi

The integration and analysis of cultural heritage metadata are very important for advancing research in the field of digital humanities.This demonstration presents two R packages, finna and finto designed to facilitate seamless access to Finnish cultural heritage metadata and ontological resources, thereby empowering researchers to conduct comprehensive analyses

within the R environment.The finna package serves as an interface to the Finna API, aggregating content from Finnish archives, libraries, and museums. It enables users to perform targeted searches, retrieve metadata, and analyze a diverse array of cultural artifacts. For instance, researchers can explore historical documents, images, and audio recordings pertinent to their studies, streamlining the data acquisition process. Complementing this, the finto package provides tools to access interoperable thesauri, ontologies, and classification schemes across various subject areas via the Finto service and its Finto AI which is an automated subject indexing service. This functionality allows researchers to incorporate standardized vocabularies into their analyses, ensuring consistency and enhancing the interoperability of their research outputs.Through this demonstration, attendees will gain insights into the capabilities of these packages, including practical examples of metadata retrieval and analysis. The session aims to showcase how finna and finto can be leveraged to enrich digital humanities research, particularly in projects involving Finnish cultural heritage materials. By integrating these tools into their workflows, researchers can enhance the depth and scope of their analyses, fostering new perspectives and discoveries in the study of the past.

Introducing the new DARIAH-Campus Content Management System

Vicky Garnett

DARIAH-EU, Ireland; vicky.garnett@dariah.eu

Since its launch in 2019, DARIAH-Campus has grown and become one of the prime destinations for reusable learning resources produced within the DARIAH ecosystem and beyond. The discovery platform now houses over 250 free, open asynchronous training and learning resources, including our own 'captured event' format, covering a broad range of digital-humanities related content including Feminism in DH, automated text recognition (ATR), performing arts, and open science.

In its initial stages, contributions were made exclusively using Markdown and the git 'commit/push' methods. This method required some existing knowledge of both Markdown syntax and the GitHub environment, or at the very least users needed to undergo a steep learning curve to get comfortable with their use. This became a barrier for many users, and also led to delays in publication of resources as errors and edits were inevitable. So, in 2021, work began with a web developer in ACDH-CH, Vienna, on implementing a Content Management System (CMS) on top of GitHub using Netlify CMS with Vercel supporting deployments and previews. The content management system proved very popular with new and existing contributors, allowing them to see in almost real-time a preview of their resource as they made edits and changes.

Nearly 4 years on, we have found that the needs of the community were no longer met, as certain features are not possible within the Netlify CMS. Therefore, we once again turned to our colleagues in ACDH-CH, Vienna, to develop a new content management system, this time using Keystatic. Keystatic is built with a modern, file-based approach that integrates seamlessly with Next.js and other modern web frameworks. It also provides a more flexible and extensible API, making it easier to customise and scale for different content management needs.

This demo will walk potential training content providers through the process of using the new content management system, with a demonstration of an example resource from the initial proposal stage up to the final publication stage. Users will also be able to make a start on their own resources, or ask questions for guidance if they are already working on a resource and need some assistance.

NewNa Segmentation App: An app to segment and dynamically interact with magazine pages

Tobias Johannes Kreten, Svend Thorbjörn Göke

Georg-August University Göttingen, Germany; tobias.kreten@stud.uni-goettingen.de, s.goeke@stud.uni-goettingen.de

This project, titled NewNa Segmentation App, evaluated the zero-shot capabilities of an existing Detectron2 model and developed an application to facilitate manual correction and expand train-ing data to improve model performance. The automatic segmentation of advertisements from historical newspapers and magazines presents a challenge for Optical Layout Recognition models trained on editorial newspaper pages. While models for this task are lacking, the Newspaper Navi-gator Model (Lee et al. 2020; Lee and Weld 2020) offers a promising approach for detecting vis-ual material in American historical newspapers. However, its effectiveness on other printed media, such as German cultural magazines from around 1900, remains uncertain. Well-segmented data is essential for further digital analyses, such as using multimodal models.

The Newspaper Navigator Model was tested on 1,789 pages from the German cultural magazine "Die Jugend." A manually annotated ground truth dataset was created and converted into COCO format for consistency. The dataset was provided by our supervisor, Johanna Störiko, who anno-tated the data as part of her PhD dissertation at the Georg-August-Universität Göttingen, based on scans from the University Library Heidelberg (https://doi.org/10.11588/diglit.3565). Initial results showed that while the model could detect advertisements, its accuracy was only around 63%, measured using Intersection over Union (IoU) with a 0.5 threshold, along with precision and re-call, from which an F1-score was derived. This relatively low accuracy highlighted the need for a tool that enables efficient correction and annotation, reducing the effort required to generate high-quality training data. To address this, we developed an interactive segmentation application inte-grating the model with a MySQL database and providing an intuitive user interface. The database schema stores annotated advertisements and complete pages. Users can modify bounding boxes, delete incorrect predictions and add new segmentations.

To enhance usability, the application includes an interactive category legend, and a toggle func-tion between modes of operation. An Upload-Only mode lets users upload pages, segment them and download results as structured JSON files and segmented images. The Database mode, de-signed for large-scale dataset creation, enables direct storage of segmented advertisements with metadata. This structured approach supports systematic data curation and model improvements. The evaluation showed that integrating machine learning predictions into an annotation tool streamlines segmentation, even when model accuracy is suboptimal. By embedding automated suggestions in a friendly interface, annotation workload is reduced while generating high-quality training data in a still human-driven interpretation.

This application demonstrates how pre-trained models can be adapted and reused in different re-search contexts. By integrating model predictions into an annotation tool, time can be saved while generating labeled data for future model training. Unlike many tools, this application is tailored for advertisement segmentation, making it highly optimized for its use case. At the same time,

its flexible architecture allows adaptation to other image segmentation tasks, provided outputs are structured in COCO format. This flexibility offers a promising avenue for further research in au-tomated document analysis and digital humanities. The NewNa Segmentation App is not just a technical innovation but a tool that enhances humanities research questions.

Towards interdisciplinary approaches to digital cultural heritage: GLAM Labs and data spaces

Alba Irollo[1], Gustavo Candela[2], Mahendra Mahey[3], Katrine Hofmann[4], Olga Holownia[5], Nele Gabriëls[6], Steven Claeyssens[7]

[1]Europeana Foundation, Netherlands; [2]University of Alicante, Spain; [3]Tallinn University, Estonia; [4]Royal Danish Library, Denmark; [5]International Internet Preservation Consortium, USA; [6]KU Leuven Libraries, Belgium; [7]National Library of the Netherlands, Netherlands; alba.irollo@europeana.eu, gcandela@ua.es, mahendra.mahey@tlu.ee, khg@kb.dk

At the intersection between the cultural heritage and research sectors, the increasing availability of digital cultural heritage data has supported new ways of producing knowledge, research, publishing results, and teaching in academic settings. The efforts put into the digitisation of their collections and capturing born digital archives, have led GLAMs (Galleries, Libraries, Archives and Museums) and other organisations such as universities to focus on many factors. These include the wider accessibility of the data and its reuse-related potential refining approaches to digital curation so that digital cultural heritage can support computational analysis, especially through Digital Humanities approaches. Emergent initiatives, such as Collections as Data, FAIR and the CARE principles for Indigenous data governance, have emphasized the need for best practice when making data available based on sets of principles (Padilla, 2023; Carroll, 2020). Moving from these principles, the International GLAM Labs Community has grown as a collaborative and interdisciplinary initiative aimed at promoting the publication, computational access and responsible use of data (Mahey, 2019; Candela, 2023; Candela, 2023b). It includes more than 90 institutions covering a wide diversity of domains from GLAMs and wider. In this context, collaborations with researchers and university support staff from different fields have become crucial factors in supporting the study of the past with a careful eye on the needs of the present and future.

- The European data space for cultural heritage fosters collaborations between the cultural heritage sector and academia, building on the experience of the Europeana Initiative and its long-standing partnerships with research infrastructures like DARIAH. These partnerships have led to joint efforts, especially in digital humanities. A new trend deserves attention in this context, and it is the increasing number of university courses focusing on digital cultural heritage across Europe. Acting as an observatory, the data space aims to foster exchanges around this topic, enhancing cooperation between the academia and cultural heritage institutions in the educational paths of the next generation of (digital) curators.

- GLAM institutions are exploring new ways to make their content available suitable for computational access. This section will introduce examples of publication and reuse of digital collections published by relevant institutions such as the Royal Danish Library and KU Leuven. It will include a description of challenges and opportunities illustrating their approach.

-The publication of digital collections suitable for computational use can be a difficult task for the institutions. Best practices and guidelines to publish digital collections suitable for computational use promote its adoption (Candela, 2023a). This section will introduce a selection of relevant and innovative initiatives that can be useful in this context. It will also provide examples of use based on a wide diversity of content such as GLAM datasets and Web Archives.

- Will highlight examples of use and lessons learned from his research with users of digital cultural heritage at the British Library and those providing services for the reuse of digital cultural heritage as data around the world in the GLAM Labs community.

Posters

Time: Thursday, 19/June/2025: 9:30am - 11:00am
Session Chair: **Alexander Steckel**, Göttingen State and University Library
Session Chair: **Stefan Buddenbohm**, Göttingen State and University Library

Workers' Voices in the Digital Age: A Newspaper-Based Digital Collection on Portuguese Self-Management Movement

João Pedro Oliveira

Faculdade de Ciências Sociais e Humanas, Universidade NOVA de Lisboa, Portugal; oliveirajoaopedro79@gmail.com

Following the Portuguese Carnation Revolution (1974), workers autonomously organised in response to the severe economic and financial crises inherited from Estado Novo dictatorship. (Fontes & Cabreira, 2020) Without trade union support, they occupied and self-managed workplaces to improve living and working conditions. (Fontes & Cabreira, 2020)

This digital collection preserves news items from two historically significant Portuguese newspapers, Diário de Lisboa and Combate. Its primary objective is to foreground workers' arguments for workplace occupations while providing insights into their organisational structures, dynamics, and fluidity. By centring workers' perspectives, the collection contributes to a more nuanced understanding of their experiences. Informed by methodologies from various Digital Humanities Conferences, the project aligns with established best practices in the field. (Terrón Quintero et al., 2023)

The collection's development follows two key methodological approaches. First, a critical analysis of newspapers extracts articles thematically relevant to the self-management movement. Second, to enhance accuracy and efficiency, PDF newspaper files are processed using the OCR tool, MasterPDF Editor, rendering documents searchable and facilitating keyword identification. Initially, 316 news items from Diário de Lisboa were compiled into a dedicated spreadsheet, which then underwent refinement to ensure consistency and correct minor syntactical errors. The final curation phase involved critically proofreading each item to maintain content quality available to the public.

This project highlights the notable lack of archival development concerning this movement. In the absence of a stable institution dedicated to preserving its memory, both academics and the public encounter significant challenges in accessing historical information. The project aims to facilitate scholarly and public engagement by developing the first fully dedicated digital collection on the *Autogestão* (Self-Management) movement in Portugal and centralising research within a unified digital platform. By encouraging users to explore the collection and conduct further investigations across archival networks and research institutions, this initiative contributes to a broader understanding of the movement's historical significance. (Sinn, 2012)

Implementing the digital collection fosters engagement with academia and the wider public. By establishing a centralised platform, this project serves as a foundational resource for researchers, ensuring accessibility to primary materials. Additionally, future collaboration with archival institutions promotes research cooperation and advancing knowledge of self-management movements. Through this initiative, users are encouraged to critically examine the materials and contribute to expanding the historiographical discourse surrounding workplace occupations. By structuring and this collection, the project enhances both academic research and public awareness, reinforcing the importance of digital humanities in historical preservation.

Of Yak Shaving and Data Taming: Building an RDF ETL Pipeline for the CLSCor Graph

Lukas Plank, **Katharina Wünsche**

Austrian Academy of Sciences, Austria; lukas.plank@oeaw.ac.at, katharina.wuensche@oeaw.ac.at

The study of Europe's literary heritage is a study of the past– of texts, traditions, and ideas that have shaped cultures over centuries. Engaging with this rich and multilingual legacy calls for methods and tools capable of navigating its complexities and interwoven histories. In the digital age, these methodological needs have given rise to innovative approaches that combine traditional humanities scholarship with computational techniques and collaborative infrastructure.

Emerging from this intersection of traditional literary studies and digital innovation, the CLS INFRA project is a joint effort dedicated to advancing the digital research infrastructure in the humanities, with a particular focus on literary and textual scholarship. Its goal is to build a shared and sustainable framework that supports literary research in the digital age.
To achieve this, CLS INFRA seeks to identify, describe, and ultimately harmonize a wide range of dedicated resources with the tools required to facilitate scholarly exploration and analysis. Central to this effort was the development of the CLSCor ontology, which was designed to serve as a standardized CRM-based ontological framework for enabling the RDF Knowledge Graph representation of a wide range of metadata properties and generally information objects in the domain of multilingual, interconnected literary heritage.
∗The term "yak shaving" is programming lingo for the seemingly endless series of small tasks that have to be completed before the next step in a project can move forward. See: https://www.techtarget.com/whatis/definition/yak-shaving 12

Building on this foundation, the process of generating the CLSCor Knowledge Graph was fundamentally organized around two high-level tasks, reflecting the inherent distinction between corpus-level and document-level metadata: 1. Collecting and persisting corpus-level metadata in a tabular data structure (referred to as 'corpusTable') and subsequent conversion of this table into CLSCor-compliant RDF. 2. Generating CLSCor-compliant RDF graphs based on corpus documentlevel metadata dynamically extracted from various sources. Expectedly, due to the highly heterogeneous nature of the available data sources, the process of CLSCor-compliant RDF generation at the documentlevel was considerably more arduous than for the single-sourced corpus-level metadata. In addition to the challenges posed by the diversity of data formats, another significant difficulty was the varied provenance of the data. Digital literary corpora and corpus documents not only exist in a wide range of formats each designed for different scholarly and computational needs but are also scattered across a fragmented landscape of storage systems and access mechanisms, each requiring a specific extraction approach. This diversity in data formats and data

provenance necessitates a flexible and sufficiently abstract approach to data extraction and transformation. Particularly, extraction components of ETL (Extract, Transform, Load) processes need to be able to not only parse various formats but also support multiple retrieval methods, from direct file access to API querying and web scraping.

To address these requirements, we conceptualized an RDF ETL pipeline capable of transforming raw data from heterogeneous sources into CLSCorcompliant RDF, validating the generated RDF against SHACL constraints to ensure structural and semantic consistency, and- upon passing data shape constraint checks- automatically ingesting the data into named graphs of a remote triplestore. Crucially, our RDF ETL pipeline was deliberately designed with a pluggable and component-driven architecture in mind. This decoupling of processing steps into modular components allowed us, for example, to overcome a key limitation of the Blazegraph triplestore, which supports either inferencing or named graphs, but not both. Since named graphs were essential for our data organization, Blazegraph's inferencing was not natively viable. By leveraging our modular design, we were readily and relatively effortlessly able to add an optional inferencing step as a component of the pipeline, maintaining named graphs without sacrificing reasoning capabilities.

Shared History, Shared Data: Unlocking World War II Victim Databases for Public Engagement

Mojca Šorn[1], Marta Rendla[1], Andrej Pančur[1], Tamara Logar[1], Vid Klopčič[2], Matevž Pesek[2], Katja Meden[1,3]
[1]Institute of Contemporary History, Slovenia; [2]Faculty of Computer Science, University of Ljubljana, Slovenia; [3]Institut Jožef Stefan, Ljubljana, Slovenia; tamara.logar@inz.si, Vid.Klopcic@fri.uni-lj.si, katja.meden@inz.si

This contribution discusses the development of the research database entitled Victims Among the Population in the Territory of the Republic of Slovenia During and Immediately After the Second World War, developed within the SIstory portal. The collection is a systematic record of military and civilian persons who had the right of residence in the present-day Republic of Slovenia during the Second World War and the immediate post-war period (May 1940 – January 1946) and lost their lives due to wartime and (revolutionary) post-war violence or the consequences of war. Currently, there is data for more than 100,000 victims, representing 6,7% of the population at the time. Each victim's identity is documented through personal data and information on the circumstances of death, comprising a total of 25 metadata fields.

The database is the result of research conducted by the Institute of Contemporary History between 1997 and 2012 as part of four major research projects. Originally, the database was designed to compile data from various historical sources while ensuring the accuracy and veracity of records through rigorous verification. However, due to the sensitive nature of the information and the ongoing war- and ideologically-charged discourse surrounding the WWII in Slovenia, only partial data was made publicly accessible in the early project phases. Specifically, details on the death of victims and status classification were collected but omitted from the publicly available records.

Recent legislative changes and the commitment to open research and public engagement prompted a shift toward greater accessibility. As a result, the database has been redesigned to not only provide unrestricted access to previously limited data, but also to enable public participation. The updated version now allows users to contribute additional information, comments and personal narratives within designated layers, promoting a more comprehensive and collaborative approach to historical documentation.

User registration is required via phone number and can then enrich the existing dataset either by correcting existing records (with support from relevant literature and sources if available) or by entering data about a new victim. The structured layers within the database ensure a clear distinction between data verified by the Institute of Contemporary History (original database) and contributions of individual users or affiliated institutions.

This project contributes to preserving historical memory and increasing transparency through open data and citizen participation. By enabling the public the opportunity to provide additional information and corrections, it promotes a more comprehensive and inclusive record. The transition to open access supports both scientific research and broader public engagement. Combining verified data with user contributions, the database provides a balanced approach to documenting a complex historical period while fostering collaboration in historical research.

Threads of the Past: Exploring Open Digital and Manually Extracted Data to Visualize Social Networks in María Lejárraga's Legacy (1874–1974)

Dolores Romero López, Patricia García Sánchez-Migallón
Universidad Complutense de Madrid, Spain; dromerol@ucm.es

From the perspective of content, this project examines María Lejárraga's contributions to feminism, modernism, and theatre during Spain's Silver Age, focusing on the social networks she engaged with and fostered from 1900 to 1936. It highlights her ability to build connections in intellectual and social spaces despite the constraints of her historical context. Lejárraga's role as a prominent yet often overshadowed author due to her husband's influence, Gregorio Martínez Sierra, is a central theme. The study concludes that her intellectual and social networks were pivotal to her emancipation and the advancement of feminism in Spain. This project offers a comparative analysis of manually extracted data from historical printed sources and open digital datasets made available by national and international institutions. By assessing both types of data, we explore their potential contributions to a social network visualization project while uncovering critical disparities in data reliability, completeness, and accessibility, emphasizing the necessity of expert-curated data for rigorous humanities research. In the specific case of this poster, the data and metadata provided by public institutions regarding the life and literary production of the Martínez Sierra spouses have been tested. These institutions include the Biblioteca Nacional de España, the Residencia de Estudiantes, the Biblioteca Virtual Miguel de Cervantes, HathiTrust, and the Revistas Culturales 2.0 Web. We also try to download data from Wikidata, but the quantity of the data was insufficient. The aim was to verify that these institutions do not provide the necessary open digital data to analyse socialization networks. We were only able to download open data from the Biblioteca Digital Mnemmosine, available on GitHub and Zenodo. Thanks to this open data and the raw data from expert bibliography (see below), we have been able to create our own database, which can serve as a foundation for future research on other authors of the so-called Silver Age of Spanish literature (1900-1939). The project also proposes a workflow that directly relies on European

infrastructures such as DARIAH. First, to find a useful social network visualization tool, we searched in the SSHOC Open MarketPlace and found Gephi as a very powerful option. On the DARIAH Campus, we found many training materials to learn how to use the tool, and after data curation and visualization creation, the resulting dataset was uploaded to the Zenodo repository and then linked back to the SSHOC Open MarketPlace. In this way, we ensure the circularity of the research, the application of FAIR principles, and our commitment to Open Science. In the future, we will create a workflow in SSHOC with these best practices, thanks to the feedback we hope to receive at the DARIAH Annual Event 2025 and publish the main results in Open Access.

A Human- and Machine-Readable Thesaurus for the Conservation of Archaeological Heritage - Development, Technical Implementation and Application in digital space

Kristina Fischer, Lasse Mempel-Länger
Leibniz-Zentrum für Archäologie (LEIZA), Germany; kristina.fischer@leiza.de

Access to the latest research on examination methods, damage phenomena, conservation techniques and materials, as well as preventive measures is essential for conservators. The evolution of conservation ethics, utilised materials and techniques over time provides valuable insights into the challenges and decisions faced by previous generations of conservators. By studying historical conservation practices, we can critically assess their long-term effects, learn from past experiences, and refine current approaches. However, a major issue in leveraging historical conservation knowledge is the scattered and inconsistent nature of the data landscape. Conservation records are stored in different formats, distributed across various databases, and described using heterogeneous terminologies. Experts from different periods and disciplines have used distinct terms for similar concepts, creating inconsistencies that hinder data retrieval, comparison, and interdisciplinary collaboration. The lack of standardised terminology not only complicates historical analysis but also impedes the development of interoperable digital resources for archaeological conservation.

To address these challenges, the "Conservation and Restoration Thesaurus for Archaeological Cultural Heritage" was developed at the Leibniz-Zentrum für Archäologie (LEIZA) as part of NFDI4Objects. One of the key objectives of this thesaurus is the systematic extraction of used terminology from digital and retro-digitised conservation reports from approximately 170 years of conservation history at LEIZA (formerly Römisch-Germanisches Zentralmuseum), supported by technical innovations like Natural Language Processing (NLP). By identifying hierarchical, equivalent, and associative relationships amongst the terms, the thesaurus enables a systematic connection between past and present knowledge, ensuring that past experiences remain accessible and reusable for future generations. This controlled vocabulary not only facilitates human communication by standardising terminology but also enables machine-readable, semantically structured data integration for the Semantic Web. Based on the Simple Knowledge Organization System (SKOS), the thesaurus ensures interoperability and follows the FAIR principles (Findable, Accessible, Interoperable, Reusable). The content classification also aligns with ISO and DIN standards.

To support the creation of this vocabulary, a user-friendly web application was developed. This application validates tabular vocabulary data against the SKOS-based schema, identifying errors like duplicate identifiers or incorrect hierarchies. It provides multiple visualisation options for hierarchical structures and enables collaborative content development through an interactive comment function. The validated data can be exported as an RDF turtle or JSON file and integrated into central thesaurus repositories such as DANTE and SkoHub or Linked Open Data platforms like Wikidata. The collaboration between domain experts and software engineers ensured both user-friendliness for professionals and machine readability for digital applications.

At the DARIAH Annual Event 2025, the methodological and technical aspects of the thesaurus development will be presented in a poster session. The presentation will highlight the practical benefits of structured vocabularies, with a particular emphasis on how standardised terminology can help unlock conservation knowledge of the past. The official release of the Conservation Thesaurus version 1.0 is anticipated before the conference, showcasing how semantic technologies can contribute to preserving and enhancing knowledge, as well as overcoming long-standing challenges in communication and data integration within the field of tangible cultural heritage.

Finding Long-Term Solutions for GRETIL, a Large Indologist Corpus

David Herting[1], José Calvo Tello[1], Maximilian Mehner[2]
[1]Georg-August-Universität Göttingen, SUB Göttingen; [2]Philipps-Universität Marburg; calvotello@sub.uni-goettingen.de

Many digital pioneers in the humanities who started in the 1990s and 2000s are now struggling to keep up with the current digital world. Not only are expectations increasing, but many projects are finding it difficult to maintain their original functionalities. After 20 years or more since their inception, the difficulties are not only technological, but also a lack of funding, diminished enthusiasm and the fact that the original leaders are no longer active and some have passed away.

One example of this is GRETIL, a collection of digital texts developed between 2001 and 2020. This resource is the largest repository of machine-readable Sanskrit texts and includes texts in other Indian languages. The corpus remains popular with scholars for quick reference and text mining and has been incorporated into several ground-breaking digital humanities projects in Indology.

Although GRETIL relied on TEI to encode its texts in the final phase of the project, the project found ad-hoc solutions for many other issues, such as its own website, its own conversion system to HTML and plain text, its own collection of secondary literature in PDF, and even its own OPAC. Not least due to its early development, at a time when most suitable e-texts were not encoded in Unicode, a major technological update was inevitable after its founder Reinhold Grünendahl retired in 2016.

In 2022, the Text+ consortium was launched as part of the German National Research Data Infrastructure (NFDI) initiative. The main objective of the consortia in this initiative is to ensure the long-term accessibility of research data, to integrate existing solutions and, in general, to improve the FAIR status of the resources. A user story by Buchholz suggested the integration of GRETIL into the Text+ portfolio. As part of the new developments of the TextGrid repository and the integration of existing corpora, we decided to publish the already converted TEI documents in this repository. We are also working on the transformation of HTML into TEI and on improving the quality of the metadata and thus its FAIR status, e.g. by using terms from

the Authority control system of the German-speaking (GND). Other components of GRETIL will be published in other repositories (eDocs and DARIAH-DE Repository).

In its environment in TextGrid, GRETIL will offer new possibilities to comfortably search and compare all texts of the collection. The keywords and categories (languages, genre, religious affiliation) now standardized will be available as individual filters for the whole corpus, allowing a more flexible filtering and querying of the corpus.

Some aspects of the GRETIL will remain as they currently are. This means that the imbalances GRETIL exhibits in certain areas, e.g. the ratio of Sanskrit to Prakrit or Tibetan texts, will carry over. However, the new environment will make it easier for new projects to expand or enrich the text material in the future, thus affording the opportunity for further revitalisation of this text corpus.

From Folklore Collections to Digital Research Infrastructures: Expanding Access, Engagement, and Analysis

Sanita Reinsone[1], Line Esborg[11], Terry Gunnell[13], Kati Kallio[2,3], Kyrre Kverndokk[5], Sandis Laime[9], Will Lamb[7], Angun Sønnesyn Olsen[5], Fredrik Skott[6], Asta Skujytė-Razmienė[8], Tim Tangherlini[12], Ida Tolgensbakk[10], Mari Väina[4], Viesturs Vēveris[1]

[1]University of Latvia (LV); [2]Finnish Literature Society; [3]University of Helsinki (FI); [4]Estonian Literary Museum (EE); [5]University of Bergen (NO); [6]Institute for Language and Folklore (SE); [7]University of Edinburgh (UK); [8]independent scholar (LT); [9]Institute of Literature, Folklore and Art of the University of Latvia (LV); [10]Norwegian Museum of Cultural History (NO); [11]University of Oslo (NO); [12]University of California (US); [13]University of Iceland (IS); sanita.reinsone@lu.lv, mari@haldjas.folklore.ee

Since the early 20th century, folklore archives have served as fundamental research infrastructures – systematically documenting oral traditions, narratives, and cultural expressions. Built through long-term collection efforts – often with public participation – these archives have been instrumental in the development of folklore studies, ethnology, and related disciplines. Despite evolving in response to disciplinary, methodological and institutional transformations, folklore archives remain vital and living repositories of cultural heritage, while the advent of digital technologies has profoundly expanded access, enhanced research potential, and reshaped their workflows and functions.

Over the past decade, large-scale digitization advancements have given rise to multifaceted digital platforms such as dúchas.ie (Ireland), Folke (Sweden) sok.folke.isof.se, garamantas.lv (Latvia), samla.no (Norway), sagnagrunnur.arnastofnun.is (Iceland), Danish Folklore Nexus scando.ist.berkeley.edu, Dutch Legend Database (Netherlands) verhalenbank.nl, kivike.kirmus.ee (Estonia), and tautosakos-rankrastynas.lt (Lithuania) exemplify this shift – providing structured access to historical and contemporary folklore materials.

Collaborative projects strengthen interoperability across linguistic and national boundaries. The ISEBEL project, an intelligent cross-collection search engine for belief legends, and FILTER, a research environment for analyzing poetic text variation, are pioneering initiatives enabling comparative folklore research across linguistic and national boundaries.

A defining feature of digital folklore archives is the integration of citizen science and crowdsourcing, enhancing accessibility while keeping folklore collections dynamic. Volunteers are involved in manuscript transcription, increasingly aided by AI-powered HTR tools, as seen in Ireland duchas.ie/en/meitheal, Latvia lfk100.garamantas.lv, and Sweden sok.folke.isof.se. Other platforms, such as minner.no (Norway) minner.no and kratt.folklore.ee (Estonia) kratt.folklore.ee, facilitate knowledge exchange. Some initiatives support artistic reinterpretation of archival content, such as “Sing with the Archives” (Latvia) dziedi.garamantas.lv and Lithuania’s “To Remember Me By” project.

A key advantage of folklore digitization is the ability to extract large-scale datasets, unlocking new possibilities for computational folkloristics. These datasets reveal structural patterns, narrative evolution, and cross-cultural connections while offering deeper insights into how folklore collections were formed and curated. E.g., FILTER project exemplifies this potential by applying computational methods to extensive Finnish-Estonian folksong corpora, while the ISEBEL project demonstrates how to enhance comparative folklore research through automated translations and metadata-driven analyses. The use of large language models in Scottish Gaelic storytelling demonstrates how synthetic text generation can expand training data for speech recognition, further supporting low-resource language technologies and computational folklore studies.

This poster presents the development of digital folklore archives as emerging digital research infrastructures – emphasizing the critical need for cross-border and multilingual integration. Emerging large language models further accelerate this shift – enabling automated processing and translation of diverse folklore corpora, including under-resourced and endangered languages. These innovations enhance usability, support archival curation and public engagement, offering new opportunities for research and international collaboration.

LLM-based geospatial data extracting: A case study based on travel literature

Dolores Sáez, Pilar Escobar, Manuel Marco-Such

University of Alicante, Spain; md.saez@ua.es, mpilar.escobar@ua.es

Cultural heritage institutions, commonly known as Galleries, Libraries, Archives, and Museums (GLAM), are exploring new ways to provide a richer experience of accessing and exploiting their collections for humanities researchers, thus encouraging their reuse not only of metadata but also of content. Several initiatives, such as Labs, are based on the creative and innovative reuse of materials published by these institutions, complying with the FAIR principles (Findable, Accessible, Interoperable, and Reusable).

The visualisation of collections within cultural heritage institutions is evolving quickly, driven by the need to improve the user experience and leverage available digital collections. Nevertheless, significant challenges remain in order to achieve easier and more efficient access to digital resources.

Artificial intelligence (AI) has emerged as a highly powerful resource in cultural heritage. Traditional Named Entity Recognition (NER) processing methods can be substantially improved through AI advances. For example, the learning process has been simplified by reducing supervised training. The innovation of this method is the application of Large Language Models (LLM) to

extract geospatial information from textual sources, thereby improving efficiency and expanding the scope of analysis in this field.

The primary objective of this study is to validate an automated system based on LLM, designed to extract geographical information from travel literary texts and subsequently render it graphically on an interactive map.

The main contributions of this work are: (a) extract the different places located in the text of the work which may be towns, cities, monuments, squares, or streets, (b) georeferencing travel literature text content through open data, and (c) visualisation on maps.

"The Atlas of the Holocaust Literature" - mapping the ghetto experience.

Kajetan Mojsak, Paweł Rams

The Institute of Literary Research, Polish Academy of Science (IBL PAN), Poland; kajetan.mojsak@ibl.waw.pl, pawel.rams@ibl.waw.pl

"The Atlas of the Holocaust Literature – Warszawa/Łódź" created and developed in the Institute of Literary Research of the Polish Academy of Sciences in Warsaw by the Department of Digital Editions and Monographs, the Research Group for Holocaust Literature under direction of professor Jacek Leociak and the scholars from the Center of Jewish Research at the University of Łódź. The project fuses such fields as digitalization, documentary work, popularisation of knowledge about the history of the ghetto, Holocaust studies, as well as urban studies.

The aim of the project is to collect written sources, describing the experience of life in ghettos, with the focus on the spatio-temporal dynamics of the events. Its principle is to use the digital possibilities to narrate the story through the prism of topography, using the form of interactive digital maps. The exact moment in time (the stage of the Holocaust) and the space (author's whereabouts) are interconnected and co-create a specific "space-time" of the texts, the topographic and chronological grid, which determines the type of experience, (and often - conditions of survival).

It is a truism to say that literary texts and memoirs are historical sources (Ankersmit 2001; Stefanowska, Sławiński 1978; White 2004). However, using both memoirs or diaries and literary works as sources of knowledge about the past poses many questions. The difficulty of working with this kind of texts as historical sources is perfectly illustrated by the work on the creation of the Atlas of Holocaust Literature.

Comparison between various testimonies created over the years (from testimonies written on a daily basis in the ghetto, to those written down decades later) sheds a light on the mechanisms of memory and their role in processing the past and creating memories.

To the problems with the mechanisms of memory and those of literary genres one should also add the issues of the emotional approach to the experienced events. This raises additional research questions for both the texts analysed and the way they are represented on the map: how to deal with emotional description of experience and transform it into the language of a map? How to incorporate this specific experience into the already existing structures of the project, where the texts developed so far are referential to the described space? We should also ask how to use this type of testimony in a digital project, whose narrative structures are sometimes not flexible enough, without losing their uniqueness.

eManuSkript: Developing Tools for Digital Manuscript Literacy

Jeremy Thompson, Mohamed Basuony

Institute for Digital Humanities, University of Göttingen, Germany; jcthomps12@gmail.com, mohamed.basuony@stud.uni-goettingen.de

As in other areas of the historical humanities, the field of medieval studies has been rapidly transformed by a growing storehouse of digitized manuscripts. Innovative research tools have not merely facilitated the study of original medieval documents, but have also brought new research questions to light. Image enhancement on photographs or scans have exposed "invisible" features latent in digital data; non-invasive image captures can document written texts trapped in book linings, delicately wrapped around saints' bones, or pasted and varnished inside of violins; script layers in palimpsests can be visually separated; manuscript fragments scattered globally across disparate repositories can be united in virtual ensembles where physical reconstitution is impossible. Although it may seem counterintuitive, this digital turn has been accompanied by a material turn, a profound reflection on the experiential and cognitive stakes of medieval media in their material reality.

In this context, the demand for manuscript literacy is arguably higher than ever––and indeed for a literacy surrounding the digital objects documenting medieval manuscripts. The eManuSkript project at the Institute for Digital Humanities, Göttingen, has received funding for two years from the Stiftung Innovation in der Hochschullehre in order to satisfy this demand. A collaborative undertaking between students, teachers, digital humanities scholars, and software programmers, the project is developing a suite of web-based apps and tutorials that will serve as teaching and study tools. This poster will present the project's plans for building a first-stop portal for studying a medieval manuscript in light of the so-called auxiliary sciences: palaeography, codicology, and bookbinding. One tool will automatically detect the elements of a manuscript's mise-en-page and allow users to extract selected visual elements to build a research corpus. A bookbinding tool will allow users to draw and describe sewing structures on a book spine. A complementary tool will enable users to build a bookbinding visually by translating SVG images of binding components alongside a guided explanation. All of these tools target students and scholars, and are being designed with both groups in mind.

In the context of this poster session, we propose to demonstrate test versions of two applications. The first is an image enhancement tool that manipulates UV/IR images or high-resolution scans to expose new visual data. The second application, a script analysis tool, enables users to measure letter strokes, angles, and distances in historical scripts and to generate cumulative statistical data about the script. With it, users can create descriptive profiles of individual letters, ligatures and abbreviations. It is aimed at advancing palaeographical studies in general and at facilitating precise descriptions of specific script samples. Both tools will help students to train their eyes and have already been tested in small classroom settings. Scholars should benefit, as well. Broader user feedback is critical at this phase since the project end date stands a year away.

We hope for a fruitful exchange at DARIAH and for constructive feedback about interface usage and user-friendliness, the desirability of supplementary features, and other conceivable learning or research goals.

Rewriting the past: A multi-faceted approach to improve quality in the NAKALA repository

Nicolas Larrousse[1], Edward Gray[2], Julie Verleyen[1], Claire Carpentier[1], Michel Jacobson[1], Hélène Jouguet[1], Sara Tandar[1]
[1]IR* Huma-Num, CNRS, France; [2]IR* Huma-Num & DARIAH ERIC; Nicolas.Larrousse@huma-num.fr

Huma-Num is a French national infrastructure dedicated to SSH (Social Science and Humanities) research projects. In order to meet one of our primary missions, that is, to provide the SSH community with solutions to preserve research data, Huma-Num has developed over the last 10 years a repository named NAKALA. The goal at the time was primarily to secure research data. In that respect, it has been a resounding success, with around 2 million files grouped in 800 000 deposits, and NAKALA has found its place in the national ecosystem.

However, the quality of the data and metadata of existing deposits is far from perfect, which is detrimental to the visibility and hence potential reusability of these datasets: but how can this broad objective of improving data and metadata quality be tackled? Given the large amount of data in NAKALA, it is clearly impossible to process each deposit individually; it was therefore decided to adopt a multi-faceted approach.

During the "Core Trust Seal" certification process, HumaNum was compelled to review a number of features associated with NAKALA. Documentation was completely revised to align with best practices and help guide users to submit better quality data and metadata, with a particular focus placed on the aspects of data preparation before depositing. To help users implement these recommendations during the process of data deposit, compliance checks were added, and autocompletion was added to several metadata fields to encourage the use of controlled vocabulary.

Three main levels of data curation were identified:

1) data whose quality is unreviewed by humans

2) data that are checked from a documentary point of view by humans

3) data already at level 2 that are also verified from a technical and archival point of view to be preserved over the long term using the platform of our partner CINES.

For levels 2 and 3, in order to check the data sets, a network of data stewards and experts was identified across France and a "moderation" workflow was created. Researchers can now reach out to local specialists to help them improve quality and receive a quality label.

The various actions described above are designed to improve the quality of future deposits of data. The remaining question is how to handle previous deposits. A study has been launched to examine the overall quality using multiple types of criteria, ranging from the content of the title and the use of URIs in appropriate metadata to more complex cross-metadata queries. The main idea is to be able to determine indicators and thus build dashboards to have a continuous vision of the global content of NAKALA and also more specifically for users in order to encourage them to improve quality.

The poster will review the results obtained by implementing these different approaches to "rewrite" the past and how these decisions impact the future developments of the NAKALA repository and the evolution of user support.

Building a FAIR Training Ecosystem for the Social Science and Humanities within the H2IOSC project

Alessia Spadi[1], Emiliano Degl'Innocenti[1], Lucia Francalanci[1], Francesca Frontini[2], Giulia Pedonese[2], Jana Striova[3], Laura Benassi[3], Antonina Chaban[3], Alessia Scognamiglio[4], Federico Boschetti[2], Pietro Restaneo[5]
[1]Opera del Vocabolario Italiano, Consiglio Nazionale delle Ricerche; [2]Istituto di Linguistica Computazionale "Antonio Zampolli", Consiglio Nazionale delle Ricerche; [3]Istituto Nazionale di Ottica, Consiglio Nazionale delle Ricerche; [4]Istituto per la Storia del Pensiero Filosofico e Scientifico Moderno, Consiglio Nazionale delle Ricerche; [5]Istituto per il Lessico Intellettuale Europeo e Storia delle Idee, Consiglio Nazionale delle Ricerche; alessia.spadi@cnr.it

The study of the past, in all its shapes, underwent a transformative evolution in the last decade given by the introduction of new technologies and digital tools to support research endeavors. Digital technologies can provide effective support in different fields in the Social Science and Humanities sector; however, the availability of these technologies does not guarantee their effective utilization. Scholars, students and researchers need training materials to engage with digital tools and methodologies to use the full potential of technology in their studies. In this proposal, the Training Environment developed within the H2IOSC project will be presented to show how it can support interdisciplinary training and continuous professional development in the Social Science and Humanities (SSH) sector.

The Humanities and Cultural Heritage Italian Open Science Cloud (H2IOSC) project aims to create a federated cluster of services and resources, developed by the Italian national nodes of four Research Infrastructures (RIs) that are part of the ESFRI (European Strategy Forum on Research Infrastructure) roadmap in the field of Social and Cultural Innovation: DARIAH.it; E-RIHS.it, CLARIN.it and OPERAS.it.

Within the H2IOSC project, the Work Package dedicated to Training, Capacity Building, Engagement aims to support research through knowledge transfer and the implementation of good practices in education. The involved RIs can share information, training and guidance initiatives that aim to promote knowledge of the products, services and opportunities offered by RIs to potential users. In the field of training, initiatives are often scattered across different platforms, not always described with specific metadata and not always accessible to the public. To address this issue, the H2IOSC training infrastructure has been developed to provide an integrated environment for accessible and reusable courses for both trainers and students, supported by the implementation of a common methodology for structuring educational materials according to the FAIR principles (Findable, Accessible, Interoperable, Reusable). The infrastructure consists of two platforms: the H2IOSC Training Environment,

which is used for the delivery and use of courses by users, and the H2IOSC Training Library, a specific repository of training materials for trainers (https://www.h2iosc.cnr.it/training-infrastructure/).

The H2IOSC Training Environment, based on a design shared by the 4 infrastructures to offer a complete experience to both students and teachers belonging to the different research areas, is a learning management system designed to offer a highly interactive virtual learning environment.

The H2IOSC Training Library is a specific repository of training materials for trainers. It is dedicated to the FAIR deposit of modular teaching materials, allowing the assignment of Persistent Identifiers (PIDs), standard licenses and integrated version update management.

The H2IOSC Training Environment and Training Library platforms aim to empower trainers and researchers in the Social Sciences and Humanities (SSH) and beyond to successfully integrate digital tools and methodologies into their work. H2IOSC is building a federated ecosystem that provides essential training resources and supports the implementation of FAIR principles in training materials as digital objects.

Needles in Haystacks? The Text+ Registry as Finding Aid for Scholarly Editions and other Resources

Daniela Monika Schulz[4,5,6], Nils Geißler[1,2,6], Kilian Erasmus Hensen[1,6], Leon Fruth[3,6], Tobias Gradl[3,6]

[1]Cologne Center for eHumanities; [2]Fachinformationsdienst Philosophie; [3]Otto-Friedrich-Universität Bamberg; [4]Herzog August Bibliothek Wolfenbüttel; [5]Universität zu Köln; [6]Text+; schulz@hab.de, nils.geissler@uni-koeln.de, kilian.hensen@uni-koeln.de, leon.fruth@uni-bamberg.de

Whilst the importance of editions for research is undeniable, their discoverability can be challenging at best. This is due to various factors: scholarly editions are usually prepared within the framework of third-party-funded projects, whereby not only the type of editions, but also the funding requirements and the formats for disseminating vary to a great extent. Databases of funding bodies such as the German Research Foundation (DFG) do contain some information on these projects, but usually lack integration with external knowledge bases. Printed editions in Germany are generally recorded in library catalogues, but it is not trivial to find them there, as no subject term is commonly used to denote them. Digital editions are for the most part not included in library catalogues at all. While there are inventories of digital editions curated by individuals or small groups, all of these services have different scopes and therefore provide different (levels of) information. Hence, users have to manually query numerous systems if they want to find all existing editions and available resources.

Within the context of the German NFDI consortium Text+, an overarching Registry has been developed to overcome these impediments. The Text+ Registry serves as a unified system to catalogue, describe and connect different types of scholarly resources (lexical resources, collections, editions), but also services, repositories and other entities such as people and institutions. The added value of the Text+ Registry lies in its overarching approach and the layering of information from different sources to provide richer descriptions. It enables researchers to conduct quick and targeted queries for relevant data, in combination with other integrated tools such as the Federated Content Search (FCS), and also provides interfaces.

Within the domain of scholarly editing, benefits compared to existing systems and catalogues result from the joint recording of both printed and digital editions, as well as completed and ongoing projects, from the inclusion of the FAIR principles in the context of digital editing , and from the integration with the basic service nfdi.software via the software registry in order to make the technical genesis of an edition transparent. The Registry can therefore fulfil a wide range of scholarly requests, such as the identification of best practice examples, or data for direct re-use (e.g., for compiling a specialised corpus), finding relevant resources for teaching purposes, or increasing the findability and thus visibility of one's own work.

This contribution presents the technologies and methods behind the Text+ Registry and discusses challenges and advantages. Its signature layering approach and the architectural design of the Text+ Registry are outlined using the domain of editions as an example.

Percy Bysshe Shelley's Influence on the British Suffrage Movement: An AI Multi-Agent system for Tracing intertextuality

Tess Dejaeghere[1], Mariaane Van Remoortel[1], Salva Ros[2], Julie Birkholz[1]

[1]GhentCDH, Ghent Center for Digital Humanities, Ghent University; [2]CLARIAH-UNED, National Distance Education University; tess.dejaeghere@ugent.be, Marianne.VanRemoortel@UGent.be, sros@scc.uned.es, Julie.Birkholz@UGent.be

Scholars have long acknowledged the radical Romantic poet Percy Bysshe Shelley (1792–1822) as a significant source of inspiration for the women's suffrage movements in early twentieth-century Britain. It is widely recognized, for instance, that the suffragette motto *"Deeds, Not Words"* was derived from his poem *The Mask of Anarchy* (1819), that Katie Gliddon clandestinely recorded her diary in a copy of Shelley's works while imprisoned in Holloway, and that numerous prominent suffrage campaigners were profoundly influenced by Shelley's revolutionary ideals and progressive conceptions of womanhood.

Building upon previous computational studies and advancing the systematic analysis of Shelley's role in the suffrage movement, we propose an agentic generative AI system designed to examine the intertextual relationships between Shelley's works and four suffrage newspapers—*Votes for Women, Suffragette, Vote,* and *Common Cause*. This approach seeks to illuminate how Shelley's poetic legacy was not merely passively received but actively reinterpreted within the discursive framework of the suffrage campaign.

This system of agents sketches a portrait of Shelley's influence on the suffrage movement, weaving together explicit references, echoes, and evolving ideas.The Mentions Agent, guardian of names that shape literary history, traces direct citations and references. The Allusions Agent reads in the shadows of the texts, uncovering ideas that slip into texts unseen. The Thematic Influence Agent follows restless concepts as they migrate, transform, and persist across time. The Citations Agent listens for the clearest echoes, tracking repeated words embedded in new contexts. Lastly, the Paraphrase Agent detects the art of saying the same thing differently, capturing how meaning reshapes itself without vanishing.

In this innovative approach, we propose a methodology that integrates expert evaluation into the assessment of system performance. This intricate task surpasses mere machine-based error counting, demanding a more nuanced and intelligent analysis—one that only human expertise can provide. Moreover, the inclusion of experts in this process will not only refine the evaluation but also foster critical discussions on the broader implications of AI-driven research in the digital humanities. assessing the viability of agentic LLM-based systems in expediting historical research

This study also advances a methodological framework that embeds expert evaluation into the assessment of system performance, recognizing that the complexity of this task exceeds the capabilities of machine-based error counting alone. A more refined, intellectually discerning analysis based on human expertise is essential. Furthermore, integrating expert oversight enhances the evaluative process and stimulates critical discourse on the broader ramifications of AI-driven research in the digital humanities, particularly in assessing the efficacy of agentic LLM-based systems in accelerating historical inquiry.

A problematic afFAIR?! Planning for the future in long-term edition projects

Daniela Monika Schulz[1,2]

[1]Arbeitsstelle "Edition der fränkischen Herrschererlasse", Universität zu Köln; [2]Herzog August Bibliothek Wolfenbüttel; schulz@hab.de

Although the role of scholarly editions for historical research is largely undisputed, as they provide the fundament for academic investigations and discourse, their preparation remains a very complex, expensive, and time-consuming endeavour. This complexity has increased even further in recent decades due to new (unforeseeable) developments and changing requirements. Traditional printed editions have long predominated and still remain popular, but the proportion of digital editions and hybrid formats is steadily increasing, as has the use of digital resources in general. (Porter 2012) Besides publishing an edition, the provision of data derived from such projects in a standardised form to make it usable for various scholarly questions, has become an eligibility criteria for funding in recent years. But what exactly is meant by ‘standardised’ form in this case, and which measures need to be taken has not yet been fully defined. The FAIR principles (Wilkinson et al. 2016) serve as guidelines, but have been formulated in rather generic terms, hence their application in the respective (disciplinary) contexts must be carefully adapted. The critical reflection on their implementation in the field of (digital) editing has only just commenced (especially in the context of the German National Research Data Infrastructure, NFDI), and remains largely a desideratum to this day. (Gengnagel et al. 2023, Hegel et al. 2023)

The ‘Edition der fränkischen Herrschererlasse’ is a long-term project funded since 2014 as part of the Academies' Programme, in which a new edition of the so-called capitularies, which are among the central legal sources of the European Middle Ages, is being prepared. The project is conceived as a hybrid edition. While the individual texts are published in a printed historical-critical edition, the accompanying digital edition also provides transcriptions of the collections as they have been preserved in the manuscripts. As a long-term endeavour, the project faces a number of challenges. One specific challenge lies in the preparation of the data in accordance with the FAIR principles. The issue of sustainability was certainly considered from the outset, but since the project started back in 2014, concrete measures for ‘FAIRification’ were not part of the original work program. (Schulz et al. 2017) How can they now be integrated retrospectively and in the most resource-efficient way? What specific measures should be implemented? What services (e.g. in the NFDI context) are available and how sustainable are they? How can overarching connectivity to other projects and databases be established in order to create added value for research?

The project is currently addressing these and other questions in addition to its day-to-day editorial work. The contribution would like to present the current considerations and concepts for discussion, but also outline the existing open questions that many (ongoing) projects are facing.

Documentation of the Polish Literary Digital Culture - Quest in the Past

Beata Koper[1], **Paulina Czwordon-Lis**[2]

[1]University of Opole, Poland; [2]Institute of Literary Research of the PAS, Poland; beata.koper@uni.opole.pl, paulina.czwordon-lis@ibl.waw.pl

The Polish Literary Bibliography (PBL) is a continuously supplemented online database collecting information on Polish literature, theater, and film history. The iPBL project, conducted within its framework, is a crucial initiative aiming at documenting online materials from the same subject area and attempting to perform a unique reconstruction of Polish literary digital culture and its history. In our poster, we would like to present the “cyber-archaeological” workflow used in the project and highlight the significant challenges of documentation of digital culture.

Digital media can hardly be called “new media” anymore: the once vibrant sites, phenomena, formats, and platforms are aging and disappearing - becoming history of utmost importance to preserve. Therefore, studying (documenting) the literary Internet requires an archaeological approach and knowledge of web history.

In Poland, there has not yet been an effort to systematically archive digital culture. There are also no relevant inventories, “maps,” or documentation that would allow starting bibliographic work in a systematic way. The poster will present different approaches to documentation: site-biographical and event-based.

The list of sources for detailed elaboration in iPBL will eventually include 200 representative born-digital magazines, services, and blogs concerning literature, theater, and film, still active on the web. The oldest compiled records (articles, literary works, reviews) date back to 1999 (1 service - from 2000, 4 - from 2001, 3 - from 2002). A list of 4,000 internet addresses of other noteworthy literary, theatrical, and film websites on the Polish web will be supplemented.

The work on the selection of sources let us isolate certain aspects of Polish literary digital culture, which flourished in the particular phases of its development in the shape of peculiar archaeological layers: some digital spaces (social chat rooms, blogs) are buried under the emerging forms of digital activity, e.g., social media.

- The phase of the pioneers of the web: when the Internet was “unfenced”, “pirate”, grassroots, open, “challenging the institutional order”, time of experimentation with e-literature, hypertexts;

- The phase of the Internet of communities: seamlessly overlapping with the previous phase, the development of “interpretive communities” and “writing communities”: forums, vortals, collaborative analysis, and literary blogs as a space of direct contact between the writer and commenting audience, situating themselves in a network of websites connected through mutual links;
- The current phase: the transfer of writers activity to social media, the commercialization of the review blogosphere, the wave of expansion and professionalization of born-digital literary magazines, new formats: podcasts, video channels, the apogee of video broadcasting of literary events during the pandemic.

In creating collections, compiling, and archiving in the iPBL digital culture project, we have encountered the following challenges:

- reaching out to as many available (though not necessarily updated) websites as possible;
- an ethical issue - selecting and describing sources: it may prove to be a gesture in the future to preserve certain sources and condemn those that have been discarded or slipped into oblivion;
- maintaining a balance between amateur and professional circuits.

Scalable refinement of the Finnish national bibliography for large-scale statistical analysis

Julia Matveeva[1], Akewak Jeba[1], Veli-Matti Pynttäri[2], Kati Launis[2], Osma Suominen[3], Leo Lahti[1]

[1]University of Turku, Finland; [2]University of Eastern Finland; [3]National Library of Finland; julia.matveeva@utu.fi

Statistical analyses of bibliographic metadata catalogs can provide quantitative insights into large scale trends and turning points in publishing patterns, enriching, and even challenging the prevailing views on the history of knowledge production (Lahti, 2019a). The use of bibliographic catalogs has become a well-established tool in literature history and helped to renew research methodology (Umerle, 2023). However, the efficient utilization of large-scale data collections as research material depends critically on our ability to critically evaluate data representativeness, completeness, quality and trustworthiness. Our earlier work has demonstrated how remarkable fractions of the bibliographic metadata curation and analysis process can be automated through dedicated bibliographic data science workflows (Lahti, 2019b, 2015; Tolonen, 2016, 2019).

This study presents further development of an open and scalable data science workflow to support literary research using the Finnish National Bibliography, Fennica. The scalability of the solutions varies by data type, and the refinement process must strike a balance between accuracy and scale. Our reproducible workflows emphasize transparency, consistency, and provenance as key elements of this process; we show how standardized refinement procedures and automated generation of versatile statistical summaries of the refined data can be used to monitor the curation process while supporting in-depth statistical analyses and modeling of publishing patterns over time and geography.

We present good practices and conceptual approaches for bibliographic data refinement and demonstrate how enriched national bibliographies can offer a data-rich perspective on Finland’s literary history. This study particularly focuses on Finland’s Grand Duchy era (1809–1917) literary analysis, contrasting manual and automated data extraction methods. The dataset, sourced from the National Library of Finland, records in Fennica from 1488 to the present day and numerous fields and subfields. We only approach around 50 fields to cater for our research needs. The workflow employs tailored methods to standardize key metadata fields, including author information, language, publisher, publication place, classification schemes, genre fields such as call number, UDC, control field and index term/genre, title, physical dimensions, and gender. These functions harmonize inconsistencies, remove ambiguities, and integrate supplementary information from external databases, ensuring high data fidelity.

The refined dataset reveals insights into Finnish publishing history, addressing gaps in metadata completeness and quality. Enrichment from external collections and complementary sources, including Finna, Kanto, and Finto, helps mitigate limitations such as missing author information, ambiguous publisher and publication place data, and the absence of gender classification. Additionally, UDC numbers were converted to words using Finto vocabulary via web scraping.

The results highlight the effectiveness of automated bibliographic data refinement in supporting large-scale research. Key outputs include a comprehensive bibliographic data science workflow, harmonized metadata dataset for research applications, and novel solutions for semi-automatic curation of national bibliographies. Informative data summaries facilitate quality control and bibliographic analysis while enabling focused studies on specific periods. The approach can be adapted for various temporal, geographic, and thematic analyses.

Dariah.hub project (2024-2025): Advancing interdisciplinary collaboration in digital humanities

Marcin Heliński[1], Aleksandra Nowak[1], Tomasz Umerle[2], Krzysztof Abramowski[1], Bartosz Szymendera[1]

[1]Poznan Supercomputing and Networking Center, Poland; [2]The Institute of Literary Research of the Polish Academy of Sciences, Poland; helin@man.poznan.pl

The Dariah.hub project (2024-2025) builds upon the earlier Dariah.lab initiative (2021-2023), which aimed to create a network of distributed digital humanities laboratories. UnlikeDariah.lab, which focused on addressing the diverse needs of researchers from variousdomains, Dariah.hub is designed to enhance collaboration through a central InterdisciplinaryResearch Platform. This new infrastructure is based on knowledge graph architecture,allowing for the integration of various disciplines such as archaeology, musicology, andsociology, while also enriching research objects.

Three levels of integration

A key aspect of the platform is its ability to integrate tools and services, including those provided by DARIAH partners, at three different levels:

1. Aggregation of research objects from digital repositories, supplying the platform with resources and data for the knowledge graph.

2. Asynchronous processing of objects retrieved from the platform using external tools that enhance them with additional metadata and resources.
3. Interactive engagement with objects through direct execution of tools from the platform, allowing researchers to work with data in real-time.

Use cases: integrating partner tools

An example of integrating DARIAH partner tools is the use of the Archaeological Module to document artifacts and archaeological sites. Once aggregated into the platform, these data can then be analyzed using historical text analysis tools, potentially provided by other partners specializing in this field.

The OCR (Optical Character Recognition) and HTR (Handwritten Text Recognition) results, obtained through tools integrated with the platform, can be further processed by NER (Named Entity Recognition) and NEL (Named Entity Linking) mechanisms, also potentially provided by DARIAH partners. This enables automatic detection and linking of semantic relationships, creating dynamic connections between entities, cultural contexts, and sociological frameworks within the knowledge graph.

Additionally, the platform supports asynchronous data processing by external interdisciplinary tools. For example, geospatial data provided by one partner could be used to contextualize archaeological findings, which are then analyzed using tools from another partner specializing in spatial analysis. This enables researchers to combine diverse datasets and methodologies, leading to a more holistic approach to analyzing source materials and uncovering previously unrecognized relationships.

Interoperability and collaboration

It is also worth noting that Dariah.hub integrates tools from the Dariah.lab suite, ensuring interoperability across disciplines. The platform offers shared workspaces with secure, collaborative editing and version control, facilitating multi-author cooperation. Automated workflows and data pipelines ensure seamless interoperability between different tools, eliminating disciplinary and institutional constraints. Additionally, open licenses support data sharing and reinforce interoperability standards.

A dynamic feedback loop between data providers and users contributes to the continuous improvement of curated datasets and the expanding knowledge graph. The platform's high-performance computing infrastructure enables resource-intensive tasks such as large-scale text corpus analysis and 3D reconstructions of archaeological sites.

Conclusion: advancing digital humanities

The poster will highlight the Dariah.hub Interdisciplinary Research Platform's role as a significant step forward in integrating digital humanities resources and tools, fostering interdisciplinary researchers collaboration, and enhancing the analysis of cultural and scientific heritage present in the knowledge graph.

Swimming in a sea of data. Digital tools for the study of Ancient Mediterranean trade and society

Manel García Sánchez[1], Arnau Lario Devesa[2], Nina Mejuto García[3], Oriol Morillas Samaniego[4], Víctor Revilla Calvo[5]

[1]University of Barcelona, Spain; [2]University of Barcelona, Spain; [3]University of Barcelona, Spain; [4]University of Barcelona, Spain; [5]University of Barcelona, Spain; arnaulario@ub.edu, nmejutga7@alumnes.ub.edu, oriolmorillas@ub.edu

Ancient societies, as any other human group, are amazingly complex and difficult to properly assess, especially when taking into account the impressive degree of cultural "globalisation" attained during the Roman empire. For this very reason, our research group focuses on several very relevant issues; by analysing inscriptions on amphorae—ancient containers used for goods like wine, fish preserves and olive oil—we can uncover vital information about ancient trade networks in the classical Greco-Roman world (5th c. BC – 3rd c. AD). Due to the scale of such endeavour, and the great amount of available data, the development of digital tools such as the CEIPAC Database of Amphora Stamps, housing tens of thousands of these inscriptions, becomes indispensable in order to identify patterns and track the movement of goods across time and space.

In addition to examining trade, the digital analysis of ancient texts and inscriptions has proven crucial for studying the representation and roles of certain marginalised sectors of society in antiquity, such as women. A digital archive compiling literary and epigraphic records of Greek and Roman women facilitates large-scale textual analysis, allowing for an exploration of gender dynamics in ancient societies. Through methods like text mining, it is possible to trace the evolution of women's representation over time and across geographical boundaries, uncovering previously overlooked aspects of social and cultural history. This digital approach not only enhances understanding of gender roles but also demonstrates the value of textual analysis in revealing new insights into the social structures of ancient civilizations.

A key objective of these digital humanities projects is to broaden access to historical research beyond academic circles into the general public, which is the one that ultimately fund such initiatives. By creating user-friendly, open access platforms that make vast databases of ancient texts and inscriptions publicly available, such as the Roman Open Data project, these initiatives invite the broader public to engage directly with historical data. Interactive platforms that host archives of ancient trade records or texts related to women in antiquity, such as the "*Gynaikes-Mulieres*" website, encourage public involvement in the process of historical discovery, thus making research accessible and engaging for diverse audiences.

By integrating digital tools like spatial analysis, machine learning, and online archives, these projects aim to redefine how historical research is conducted and shared. Whether through the digital reconstruction of trade networks using Network Science, the analysis of gender roles giving agency to ancient women, or the creation of interactive platforms for public engagement, these efforts reflect the growing impact of digital humanities on the study and presentation of the past. They also underscore the potential of digital technologies to engage new audiences, enhance scholarly analysis, and open up new avenues for interdisciplinary collaboration, advancing both research and public engagement with the ancient world.

Systematic Research Data Management at the Göttingen Campus - Showcasing the National Research Data Infrastructure

Stefan Buddenbohm, Alexander Steckel, Lukas Weimer

Göttingen State and University Library, Germany; buddenbohm@sub.uni-goettingen.de, steckel@sub.uni-goettingen.de, weimer@sub.uni-goettingen.de

Göttingen: A Strong Player in the National Research Data Infrastructure (NFDI)

This poster introduces all NFDI consortia represented at the Göttingen Campus and showcases the broad disciplinary spectrum of research data management.

Various institutions of the Göttingen Campus are involved in 17 out of the 27 NFDI consortia. The strategy for creating local support structures is increasingly driven by the Göttingen State and University Library (SUB) and the GWDG, the datacenter for the Georg-August-Universität Göttingen and the Max Planck Gesellschaft.

What is the NFDI?

NFDI systematically indexes and networks valuable scientific data for the entire German scientific system and makes it available for sustainable use. Until now, such efforts to provide sustainable data access have mostly been pursued on a decentralised, project-related or temporary basis.

NFDI represents Germany as a mandated member of the European Open Science Cloud (EOSC). NFDI is also a member of the internationally active Research Data Alliance (RDA).

The NFDI consortia with Göttingen participation cover a broad range of disciplines, thereby emphasizing the campus' ambition to deliver outstanding achievements across a wide disciplinary spectrum. Notably, the involvement in all four funded consortia for the humanities and cultural sciences highlights Göttingen's long-standing tradition of expertise in the necessary infrastructure for these fields.

- Base4NFDI (core services for the NFDI; all funded NFDI consortia from all three funding rounds are involved, with the University of Göttingen taking on responsibility in governance as a co-applicant)
- DAPHNE4NFDI (Data from Photon and Neutron Experiments for the NFDI, involvement as co-applicant)
- FAIRAgro (agroecosystem research; the Department of Crop Sciences at the University of Göttingen, involvement as participant)
- FAIRmat (FAIR Data Infrastructure for Condensed-Matter Physics and the Chemical Physics of Solids, involvement as participant)
- KonsortSWD (consortium for the Social, Behavioral, Educational and Economic Sciences, involvement as participant)
- NFDI4Biodiversity (Biodiversity and Environmental Data, GWDG as co-applicant)
- NFDI4BIOIMAGE (research data management for microscopy and bioimage analysis; the Excellence Cluster "Multiscale Bioimaging" at the University of Göttingen is involved as a participant)
- NFDI4Chem (Chemistry Consortium in the NFDI, involvement as participant)
- NFDI4Culture (Consortium for Research Data on Tangible and Intangible Cultural assets, involvement as participant)
- NFDI4Earth (Adresses Digital Needs of Earth System Sciences, involvement as participant)
- NFDI4Energy (interdisciplinary energy systems research; the Sociological Research Institute Göttingen (SOFI) e.V. is involved as a co-applicant)
- NFDI4Health (National Research Data Infrastructure for Personal Health Data, Universitätsmedizin as co-applicant)
- NFDI4Ing (National Research Data Infrastructure for Engineering Sciences, GWDG as participant)
- NFDI4Memory (research data management for historical data; the Academy of Sciences in Göttingen, the SUB, and the GBV central office (VZG) are involved as participants)
- NFDI4Objects (research data infrastructure for the material heritage of human history; the GBV central office (VZG) is involved as a co-applicant, and the SUB as a participant)
- NFDIxCS (research data infrastructure for computer science; the GWDG is involved as a co-applicant)
- Punch4NFDI (Consortium of Particle, Astro-, Astroparticle, Hadron and Nuclear Physics, university as co-applicant)
- Text+ (Text- and Language-Based Research Data, university as co-applicant)

Increasing the discoverability of research services and resources through contextualization and community use cases in the SSH Open Marketplace

Stefan Buddenbohm[2], Edward J. Gray[1], Cristina Grisot[3], Michael Kurzmeier[1]

[1]DARIAH-EU, Germany; [2]Göttingen State and University Library; [3]Swiss National Data and Service Center for the Humanities; sbudden@gwdg.de, edward.gray@dariah.eu, cristina.grisot@dasch.swiss, michael.kurzmeier@dariah.eu

Introduction

The Social Sciences and Humanities Open Marketplace (SSHOMP) is a discovery portal for Social Sciences and Humanities research communities. It showcases solutions and research practices for the research data life cycle and facilitates

discoverability and findability of resources that are essential to enable sharing and re-use of workflows and methodologies. With a population of ~5000 items, the SSHOMP relies on community curation to ensure the catalogue remains up-to-date and useful for researchers. Curation routines, mixing automatic and manual tasks, are set up to ensure and continuously improve (meta)data quality.

Contextualization and use cases

Contextualization is one of the key pillars of the SSHOMP (Barbot et al. 2024). It is meant to provide a discovery portal for tools and services, while placing these tools and services in context via publications, training materials, datasets, and workflows. As such, these last four categories are indexed in the SSHOMP insofar as they can be placed in relation with tools and services. This is an objective we are pursuing, through the automatic creation of relations and their manual curation, and through encouraging authors to create relations to other items when they create new items. This poster shows how the SSHOMP facilitates diverse methods of studying the past via contextualization of resources, relying on three community use cases:

- The integration of items created within the scope of the ATRIUM project
- The integration of items created within Text+ and DARIAH-DE
- The integration potential of items, including workflows, originating from a DARIAH WG dealing with historical data

ATRIUM Project

The network analysis in our poster shows how the SSHOMP provides insights into the use of tools, methods and standards in the DH research communities, and how it increases serendipity in the discovery of new methods and standards, by interlinking the resources and describing workflows. These relations demonstrate how inter-related this specific catalogue of tools is with the overall catalogue, and the broad impact that initiatives like ATRIUM can have on the community.

Text + and DARIAH-DE

Text+ along with the Society for Humanities and Cultural Research (GKFI), use the SSHOMP as an aggregator and delivery service to present their offerings. Much like ATRIUM, resources are tagged in the Marketplace with minimal metadata and harvested regularly via the API, allowing for the portals of Text+ and GKFI to display over 80 services on institutional websites with minimal effort - needing only to implement harvesting and display: creation and curation of resources are managed solely through the Marketplace, which is a huge benefit for both entities.

DARIAH WGs

Currently, DARIAH has four WG groups focusing on historical data: ARChitectural HEritage Thesaurus through Integrated digital Procedures and Open data (ARCHETIPO), Digital Practices for the Study of Urban Heritage (UDigiSH), Digital Numismatics and Women Writers in History. Through their cross-country and cross-disciplinary character, these WGs create unique resources, tools and knowledge about the past. The SSHOMP is a powerful tool for the dissemination of these resources, and for translating their knowledge and expertise into step-by-step, practical workflows.

Reconstructing urban transformations: Digital Humanities for the documentation of large-scale construction sites in historic cities

Sofia Darbesio

Politecnico di Torino, Italy; sofia.darbesio@polito.it

Large-scale architectural construction sites are a relevant part of cities' contemporary developments. They also produce a high amount of diversified data, tools, and information that is complex and challenging to access or communicate. However, especially in the case of historic cities, they embody inventive capabilities, innovation, and processes, representing relevant information for architectural and urban history. By applying Digital Humanities approaches, it is possible to explore new ways of documenting, studying and narrating inherent complexities such as decision-making processes, the interaction of interdisciplinary urban history actors, and the contextualisation of spatial-heritage relationships, while framing these dynamics in the context of the historic city and its past.

This research addresses the challenges of recording and preserving construction site documentation by producing a critically structured digital library of diverse (born-digital, digitised and non-digitised) data, metadata and resources. By intersecting historical and contemporary materials through ICTs, a new interactive multimedia solution can provide a dynamic virtual representation of the urban space and its past. By representing site processes through a spatialised digital reconstruction of phases and interactions, the system can document and interpret the evolution of architectural worksites in historic cities, offering transparency and insight into the spatial-cultural relationships that shaped the present identity of the urban space. Therefore, the research deals with producing a digital prototype to allow cross-referencing the current and past versions of the site-related materials, promoting the accessibility and sustainability of the collected information. This digital interface will make it possible to visually communicate the dynamics of urban development through different historical-critical narratives.

The chosen demonstrator is the Piazza Municipio metro station in Naples, a large-scale infrastructural and architectural worksite in the historic city centre that has taken over twenty years to complete. Designed in 2003 by the Portuguese Pritzker Prize-winning architects Álvaro Siza and Eduardo Souto de Moura, the architectural project dealt with multiple historical heritage segments. During the excavation phases, important archaeological evidence emerged, testifying to the overlapping of multiple historical layers throughout time due to the succession of many different cultures and social changes. These findings added complexity and depth to the worksite process, making it even richer and more multifaceted.

By addressing these dynamics through a Digital Humanities approach, the study contributes to the understanding of large-scale urban transformations and their relationship to urban history, providing new digital means to document and narrate the city and its past. In this context, this contribution focuses on the challenges of building a digital tool to reconstruct, make visible, map and narrate complex processes of a large-scale contemporary construction site as a spatial and conceptual node of the historic city at the crossroads of its past, thus leading to a more comprehensive understanding of its connection to urban history.

AI-Enabled Citizen Participation in Safeguarding Ukrainian Cultural Heritage: Ethical and Methodological Frameworks

Tugce Karatas[1], Sanita Reinsone[2], Marianna Ziku[5], Uldis Zariņš[2], Katerina Zourou[5], Pavlo Shydlovskyi[3], Alba Irollo[4]
[1]University of Luxembourg, Luxembourg; [2]University of Latvia; [3]Taras Shevchenko National University of Kyiv; [4]Europeana; [5]Web2Learn; tugce.karatas@uni.lu

The preservation of Ukrainian cultural heritage faces unprecedented threats due to ongoing geopolitical turmoil. Destruction, displacement, looting and loss of cultural artifacts require urgent and innovative responses to safeguard both tangible and intangible heritage. Artificial Intelligence (AI) presents new opportunities for digital preservation, documentation, and restoration, yet its application raises ethical and methodological challenges. This poster explores how AI, when combined with citizen participation, can be effectively and responsibly leveraged to protect Ukrainian cultural heritage, ensuring an ethical, sustainable, and community-driven approach to digital preservation.

The AISTER project, funded under the Erasmus+ KA2 programme, employs AI technologies alongside active citizen engagement to develop participatory models of cultural heritage safeguarding. It aims to advance AI-driven methodologies, including multilingual text recognition, AI-powered image analysis, and 3D reconstruction of cultural sites. These technologies facilitate the documentation, restoration, and digital preservation of endangered heritage assets while actively involving communities in the process. The project aligns with European Union regulations and UNESCO guidelines, ensuring AI applications uphold ethical principles related to transparency, inclusivity, and sustainability. Responsible AI use is crucial to mitigate biases, prevent the misuse of sensitive cultural data, and support heritage professionals in navigating AI-driven decision-making processes. Beyond technological innovation, AISTER underscores the importance of community engagement and public awareness. The project fosters participatory approaches where university students, cultural heritage professionals, and citizens collaborate to identify risks, document artifacts, and develop AI-enhanced heritage preservation strategies. Through co-creation workshops, hackathons, and roundtables, AISTER promotes knowledge exchange and strengthens public involvement in heritage safeguarding. These activities help demystify AI, making it more accessible and fostering public trust in AI-assisted preservation efforts. The project's outcomes will contribute to ongoing discussions on AI ethics, human-centered AI, and citizen participation in cultural heritage preservation. In addition to expert roundtables and open-access research publications, AISTER will organise an "AI for 3D Cultural Heritage of Ukraine" hackathon, community-driven AI workshops, and a structured policy framework for ethical AI use in the heritage sector. The AISTER Manifesto will further outline best practices for integrating AI in heritage protection, emphasising ethical considerations, public engagement, and long-term sustainability.

By fostering interdisciplinary collaboration between computer science, cultural heritage, and humanities and social sciences, AISTER showcases AI's potential beyond documentation and analysis, serving as a catalyst for engagement, awareness, and collective responsibility in the preservation of cultural heritage. This poster aims to contribute to broader discussions at the DARIAH Annual Event, highlighting the intersection of AI, digital humanities, and participatory heritage preservation to ensure an ethical and sustainable approach to digital humanities research.

An Open Access database for Khmer Buddhism (Cambodia): enhancing iconography with Omeka-S

Juliette Lecorney[1], Lauriane Locatelli[2]
[1]University of Strasbourg, France / DISTAM; [2]INIST - CNRS; lecorney.juliette@gmail.com

The aim of this poster is to present an iconographic database gathering Buddhist images from Ancient Cambodia. As a matter of fact, the purpose of my doctoral dissertation is to study the specific features and evolution of Buddhism in Ancient Cambodia (pre-angkorian and angkorian periods). In this context, iconography is one of the main sources. However, ancient Khmer Buddhist images are plentiful and scattered in collections around the world. Thus, one part of my doctoral research is to create a database to collect and centralise Buddhist images from Ancient Cambodia. The objective is also to encourage and simplify cross-referencing between sources, using mapping tools and links with epigraphic data. The aim of this iconographic database (containing statues in particular) is to promote and make available to the scientific community and the general public, iconography from the pre-Angkorian and Angkorian periods (from the earliest attestations to the reign of Jayavarman VII, 1181 - approx. 1218) in the Far East (Cambodia, Thailand, Laos and Vietnam).

This website and database is developed with support from INIST-CNRS and the DISTAM Consortium. This project is aligned with the objectives of Open Science and FAIR data. The database mostly featured photographs taken during my research missions in Cambodia, but also images from museums and institutions. This is a multidisciplinary approach, combining the history of religions, the study of written sources and iconography. This work is part of the Open Science approach through its compliance with FAIR principles and the digital humanities aspect, particularly as regards cartography with GeoNames.

Thus, one part of my poster will focus on the methodology, the technical aspects and the construction of the various tools (cartography, search tool, classification, etc.). The other part will focus on the benefits offered by the creation and the use of such a database in the field of Buddhist (and Khmer) studies. Thereby, the aim of this poster is to present this database and its tools specially designed for Buddhist studies. But it will also provide an opportunity to discuss on improvements and additions that can be made, as well as the limits of this project. Finally, this poster will also raise questions relating to image licences and the use of images by scholars.

Aspect Detection and Classification in Historical Travel Literature: A study on Prompting Strategies and on the Diachronic influence of Language on Generative AI Performance

Tess Dejaeghere[1], Salvador Ros[2], Julie Birkholz[1]
[1]GhentCDH, Ghent Center for Digital Humanities, Ghent Universityn; [2]CLARIAH-UNED, National Distance Education University; tess.dejaeghere@ugent.be, sros@scc.uned.es, Julie.Birkholz@UGent.be

In Digital Humanities (DH), the recognition, extraction, detection, and classification of aspects are fundamental for tasks such as entity linking and network visualization. Traditionally, these tasks have relied on rule-based methods or discriminative language models functioning as classifiers, (Dejaeghere and Singh 2024). However, the rapid development of Generative AI models, including LLMS as GPT-4, Gemini, Llama 3, and Claude, (Openai 2024; Google 2024; Anthropic 2023), presents a significant opportunity for advancing aspect task. These models enable users to engage with extensive training data through natural language instructions, reducing the need for manual feature engineering or large annotated datasets. The increasing accessibility of chat-based interfaces, (Amazon 2023). further lowers the technical barriers for researchers and practitioners in DH, facilitating experimentation with information extraction techniques.

However, scholars require clear guidelines to effectively utilize LLMs for information extraction. To achieve this, it is essential not only to conduct a systematic study of prompting techniques and their impact on extraction performance but also to analyze the diachronic effect of language on model accuracy, as this remains a critical research challenge. Therefore, it is necessary to answer these questions:

1.- Which prompting strategies most successfully detect and categorize aspects?

2.- Do language models perform better on texts from more recent centuries, where language is more standardized, or do they perform equally well on texts from earlier periods?

3.- Do language models predispose to perform better or worse when detecting different categories of aspects?

For this purpose, we examine the effectiveness of Generative AI models and prompting techniques for aspect recognition and classification in historical travel literature, focusing on English texts from the 18th, 19th, and 20th centuries (Dejaeghere and Singh 2024). Using an annotated dataset containing entities related to travelers' environments—such as fauna, flora, weather, locations, and organizations— we leverage different LLMS (e.g. LLama 3.2, GPT4o) to evaluate four prompting techniques, across different time periods and entity categories. A category-specific ablation study assessed the impact of both prompt design and linguistic variations across centuries on aspect detection performance. This approach provided deeper insights into the interaction between historical language variation and AI-based extraction techniques.

Statistical analysis, including Kruskal-Wallis and Mann-Whitney U tests, revealed that tailoring prompting strategies to the specific objectives of the aspect detection task is essential. Among the tested prompting techniques, few-shot prompting and chain-of-thought (CoT) prompting yielded the highest precision. However, these methods did not significantly outperform others, suggesting that the trade-offs between precision and recall should be carefully considered based on the task's requirements. When maximizing the number of detected aspects, zero-shot and CoT prompting were more effective, though they required additional validation to ensure completeness. In addition, the study further highlights variability in aspect detection performance across linguistic periods, with texts from the 18th and 19th centuries outperforming those from the 20th century. This finding underscores the impact of training data alignment with historical linguistic characteristics. Future research will expand this analysis by testing other high-performing generative models, assessing their cross-linguistic and cross-model generalizability.

Identification of Coptic Dialects Using Supervised Machine Learning

Peter Missael

University of Göttingen, Germany; peter.missael@stud.uni-goettingen.de

Dialect identification plays a crucial role in understanding the linguistic and cultural nuances of the Coptic language, the last stage of Ancient Egyptian (an Afro-Asiatic language). Despite its historical significance, there has been limited research in this area. Most machine learning models focus on only one or two dialects (cf. Smith and Hulden 2016; Zeldes and Schroeder 2016; Levine et al. 2024), as these two comprise the bulk of Coptic texts. This study presents a machine learning model for dialect identification of the Coptic language, addressing the existing gap in linguistic research. Using supervised machine learning in dialect identification for other languages showed promising results (cf. Doostmohammadi and Nassajian 2019; Jauhiainen et al. 2022; Vaidya and Kane 2023).

Various methods were evaluated, including Support Vector Machine (SVM), Random Forest Classifier, Multinomial Naïve Bayes (NB), Logistic Regression, and Recurrent Neural Network (RNN) with a Long Short-Term Memory (LSTM) layer. The best performing method was Multinomial NB with an F1-score of 0.92, while most methods achieved an F1-score of 0.91.

The dataset comprises texts from six (sub)dialects, written in Coptic Unicode. Preprocessing involved removing diacritics and punctuation marks, and splitting texts into sentences, each labeled by dialect.

Feature extraction was performed using TF-IDF 1- to 2-grams. Grid search cross-validation (cv=5) was used to identify the optimal parameters for each method. The imbalance in the dataset impacted the results, as shown in the confusion matrices, with the three most represented dialects being the most accurately identified.

However, there is room for improvement. The accuracy of identifying the underrepresented dialects can be improved through digitizing more texts from these dialects. The model can be fine-tuned to identify additional dialects and enhance the identification of existing ones. It can serve as an initial step in a Coptic NLP pipeline or in research to identify the most characteristic features (words) in each dialect, particularly in texts which show influences from various dialects.

Teaching late antique and byzantine illuminated manuscripts through digital humanities. A field report

Thorben Langer, **Johanna Störiko**

Georg-August-Universität Göttingen, Germany; thorben.langer@uni-goettingen.de, johanna.stoeriko@stud.uni-goettingen.de

The complex demands placed on teaching in the field of Digital Humanities become immediately apparent when considering the students enrolled in a typical course in Göttingen. For our practical exercise, "Digital Analysis of Illuminations in Late Antique and Byzantine Manuscripts," we had 24 students enrolled in the winter semester 2024/25. Of these, approximately half (13) were pursuing a Master's degree in Digital Humanities, while 8 students were studying Data Science or Computer Science, and a few came from Iranian Studies or History. Only a few students had prior experience with historical manuscripts, and even fewer had any familiarity with the Late Antique or Byzantine periods. The three manuscripts (Madrid Skylitzes, Ashburnham Pentateuch,

and Mutinensis graecus 122) that we prepared as examples for the course were unknown to the students at the outset. Therefore, we faced the challenging task of not only teaching theoretical and methodological competencies in Digital Humanities to a group with such diverse skills and fields of study, but also of introducing them to the scholarly motivations of Late Antique and Byzantine archaeology and art history.

But how can the balance between technical processing and historical interpretation be achieved in practice in the classroom? In this experience report, we would like to share our observations on this matter. We will present the structure and content of our course and reflect on the difficulties and learning effects experienced by the students.

The core objective of the course was to teach students digital methods for analyzing illuminations in manuscripts. We aimed to introduce both established and experimental methods from Digital Image Science and Computer Vision (Image Captioning, SVG-Annotation, Image Clustering, Shape Analysis, Image Segmentation, Face Recognition). As a foundation, the students received an introduction to the various illuminations and a basic historical contextualization of the codices. Additionally, the students were encouraged to explore the codices on their own.

Following this, the digital work with the manuscripts began. We would particularly like to highlight the work on the Madrid Skylitzes. Initially, the students annotated all the heads of people in the illuminations using SVG polygons. This resulted in a dataset of over 2500 heads. We then statistically analyzed this dataset and performed a clustering of the heads based on image embeddings. It became apparent that the students not only needed practice and guidance in applying the tools, but also in fundamental image-scientific methods such as describing or categorizing images.

We therefore believe it is essential that these competencies are not taken for granted in DH teaching, but taught actively. Only then can the diverse backgrounds of the students be taken into account, and an exchange on equal terms be facilitated. This is particularly relevant for students of Computer Science or Data Science. Despite their great interest in the humanities, they often face significant challenges, while their expertise can greatly enrich the work of students with a humanities focus.

Enhancing Historical Learning Through Digital Tools: A Wikipedia-Based Teaching Innovation in Archaeology

Jordi Martín i Pons[1,2]

[1]Universitat de Barcelona; [2]CEIPAC (Center for the Study of Provincial Interdependence in Classical Antiquity) University of Barcelona; jmartinpons@ub.edu

Objectives and Theme

This project is part of an innovative teaching initiative in the Historical Sources course of the Archaeology degree at the University of Barcelona. In this course, students are invited to create a Wikipedia page, in open access. The proposal aims to leverage digital tools to enrich the teaching and learning of the past through free, open-access resources.

Teaching Innovation and Digital Humanities

The project's primary objective is to promote a better understanding of the past by applying innovative digital methodologies. Students are at the center of learning and content creation, with the guidance of an instructor. The focus is not only on studying ancient cultures but also on utilizing digital technologies to enhance history teaching and transmission through accessible media. Using Wikipedia as a platform provides an effective way to present historical information, creating significant educational and social impact.

A key part of the project is analyzing and comparing the average length of articles in languages with many speakers (e.g., English, French, Russian) versus those with fewer speakers (e.g., Basque, Catalan, Finnish). The project assesses the impact in terms of visit numbers and article quality, using sources and internal quality assessments. This allows students to observe the impact of their article, its role in helping speakers of their language access information, and its connection with other Wikipedia entries tracked through metrics.

Benefits for Society and Academic Research

Through this activity, students learn not only about historical sources but also engage in creating digital knowledge for the academic community and the public. They are introduced to tools like Wikipedia and made aware of its mission (free and cooperative knowledge). Students also learn citation practices and how to use images correctly, enhancing their understanding of digital technologies and historical data analysis. This practice bridges the gap between digital tools and historical research, enabling students to create a digital narrative of the past and understand its impact.

Open Access and Minority Languages

Additionally, creating content in Catalan demonstrates support for minority languages in the digital space. The availability of historical information in languages like Catalan is crucial for ensuring their survival. This initiative addresses the need to preserve linguistic diversity in the digital age, a critical challenge today. The project has both academic value and social benefits, facilitating access to quality knowledge in a language with less global presence. This example can inspire other scholars to apply similar strategies to underrepresented languages, improving their visibility online in alignment with the European Charter for Regional or Minority Languages.

Academically, this activity promotes a methodology combining primary research, digitization, and the continuous creation of editable, accessible content. It fosters an active learning process that empowers students to contribute meaningfully to global digital knowledge creation.

This proposal stands out for linking traditional teaching with digital humanities. Through this, students not only gain expertise in analyzing historical sources but also contribute to creating and disseminating historical knowledge in the global digital space. The project has resulted in significant satisfaction, reflected in students' demonstrated interest.

Cultural Data in Australian History: An Intimate Analytics Methodology

Rachel Fensham[1], **Tyne Sumner**[2], **Nat Cutter**[1]

[1]University of Melbourne, Australia; [2]Australian National University; rfensham@unimelb.edu.au

Cultural Data in Australian History: An Intimate Analytics Methodology

This poster begins by defining cultural data and identifies 5 Australian digital collections that curate cultural data; as such, they are knowledge structures that narrate the past, as Jurissi Parikka with his concept of media archaeology argued. It examines these extensive curated databases for the performing arts, architecture and visual arts history (https://www.daao.org.au/;https://www.ausstage.edu.au/pages/browse/; https://qldarch.net/; https://www.womenaustralia.info/; https://researchdata.edu.au/circus-oz-living-archive-collection/939530), in order to find aggregated and comparative approaches to cultural analysis.

Cognisant of Australia's settler-colonial history, the conceptual framing of the research infrastructure project, the Australian Cultural Data Engine (ACD-E, 2021-2023), developed a cross-walk architecture, that retained unique entitites while facilitating a more critical analysis, in ways that expand the logics of the originating research communities. Accessing the affordances of interoperability– generated in the friction that exists between data custodians and data engineers – we identified rich deposits of historical insight about art, artists and cultural change.

With the case study of *Know My Name*, a major national exhibition as a discrete dataset, the project proposes an intimate analytics methodology that harnesses the database as well as a close historical reading of contexts that shape the data. Following, for instance, the question of careers in the arts, we examine event datapoints that accumulate over time, as well as consider the extent to which key markers of success impact on the trajectory of a career. Moreover, being able to aggregate diverse datapoints via mapping technologies, we identify how artistic networks and scenes were indicative of other longer-term social formations.

Recognising that the database is always incomplete, misleading, and sometimes violently empty, we exemplify at a granular level how the complexity of relations between data and place, dataset and person, database and narrative must be reconceived, particularly when indigenous knowledge traditions relating to place or naming conventions are articulated. When pursued with vigour, this approach, that we term *intimate analytics*, can enhance, trouble, and unsettle conventional approaches to cultural data research, and put it into dialogue with the existing structures, biases and conventions of knowing the past.

Subsequently, we argue that such born-digital cultural collections must grapple not only with their distinctive content but also the historical contexts embedded in the cultural data itself. As repatriation and decolonial curation efforts become increasingly prominent in Europe and North America, our physical location in the Global South represents an imperative and an opportunity to speak from the 'periphery' to the 'centre' of global cultural data infrastructures and digital heritage.

Pervisum: a Tool for Digital Storytelling and Writing on the past in scholarly publications

Bulle Tuil Leonetti[1], Margaux Faure[2]

[1]INVISU (CNRS/INHA, France); [2]INHA (France); bulle.leonetti@inha.fr, margaux.faure@inha.fr

Led by the InVisu research unit (CNRS) and the Digital Research Department of the French National Institute of Art History (INHA), the PerVisum project is funded by the National Fund for Open Science (COSO). It was developed in response to a growing demand among scholars and the GLAM sector for the ability to construct scientific narratives about the past. The challenge was to leverage the increasing availability of digital images in open access while maintaining digital sobriety, a goal that the use of IIIF technologies could facilitate.

Furthermore, images used in support of scholarly publications have not yet fully benefited from the evolution of the Web. Most of the time, we limit ourselves to simply providing access to images in regards of texts, thereby missing the opportunity for real integration into the research workflow.

With the Pervisum project and tool, relying on IIIF technologies, we can integrate functionalities such as annotation, in-depth exploration of images and persistent links to sources. For instance, when illustrating the evolution of an urban landscape, the work of a painter among his peers, or the coinage of a bygone empire, scholars can now employ a vast collection of curated images interwoven with text.

The Pervisum project explores the potential for publishing scientific demonstrations as IIIF manifests, rethinking the relationship between textual and visual content. The annotations made on IIIF images form the core of the scientific argumentation.

How does it work?

The tool provides interfaces that enable users to construct demonstrations based on IIIF manifests published by heritage institutions. Users can select images and organise them according to their text plan, integrating them into the demonstration. The process involves annotating the images to support the argumentation. The writing of the demonstration is therefore based on the idea that the argumentation is contained in IIIF annotations, which are defined and ordered by the user according to the framework they wish to set up.

The objective of this poster is twofold: firstly, to present the tool currently under development, and secondly, to consider IIIF manifests as editorial objects that can be used both to disseminate enriched image corpora in publications and to offer a new article format.

A progress report of the Corpus Musicae Ottomanicae on the challenges of data modelling of historical Middle Eastern music manuscripts

Sven Gronemeyer[1,2]

[1]Max Weber Stiftung, Germany; [2]La Trobe University Melbourne, Australia; gronemeyer@maxweberstiftung.de

The Corpus Musicae Ottomanicae (CMO) is a long-term research project focusing on nineteenth-century Ottoman music manuscripts and their critical edition. Many of these manuscripts are written in Hampartsum notation and later in Western staff

notation. CMO has collected more than 10,000 musical sources and expressions and edited more than 550 pieces over the past nine years.

The musical pieces are historically and textually accurate transcriptions, reflecting the original sources as closely as possible. Similarly, the sung poetry follows the original arrangement and orthography of the sources as faithfully as possible. The musical transcriptions are fully digitized, considering the special notational requirements of Middle Eastern and Eastern Mediterranean music. The editions will be stored in a variety of formats in order to provide access to the edition in both graphical (human readable) and semantic (machine readable) form.

The latter aspect in particular is a challenging undertaking and will be the focus of this presentation. The data modelling for the editions was (and still is) in an area of conflict between the schemas of existing and established formats (namely TEI and MEI) and the editorial requirements. Here, particularly in the case of the MEI, one can see a certain Western bias in the use or definition of certain elements and schemas, based on the source materials on which they were modelled. CMO is an example of how creative use of the existing guidelines can not only meet the needs of the project, but can also stimulate discussions on how to adapt the guidelines and, in particular, open them up to perspectives from non-Western contexts.

While Western staff notation is strict and precise in terms of meters, pitches and durations, early Hampartsum notation is characterized by a minimal stock of durational signs with relative values, but differentiated pitch signs. The system later underwent a specification of durational values and a reduction in the number of pitch signs. In the 20th century, leading Turkish musicologists tried to create pitch systems that could represent Ottoman music in the most rational and efficient way. Transcription into Western staff notation is therefore always time-specific, and the original notation must be correlated with it. This was eventually achieved through the definition of custom symbols using the semiotic triangle, where the original notation could be correlated with both the Ottoman pitch names and the Western pitches. These relationships may change as a result of ongoing research, so a central reference file would be ideal. Yet, MEI does not allow for linking custom symbols in the appropriate part of the header – hopefully stipulating a change in a future guideline version.

This example shows that the maintenance of formats and standards can benefit from input from ongoing research projects. Equally, research projects can benefit from the dissemination and exchange of best practice. With this presentation, CMO aims not only to provide an overview of its efforts to make a historical music tradition available, but also to advocate for more networking to improve technical guidelines for digital scholarly editions.

Reviewers

Georgios Artopoulos
Anne Baillot
Florian Barth
Agiatis Benardou
Megan Black
Jan Brase
Stefan Buddenbohm
Costis Dallas
Caleb Derven
Vicky Dritsou
Kim Ferguson
Stefan E. Funk
Vicky Garnett
Rita Gautschy
Francesco Gelati
Elena Gigliarelli
Mathias Göbel
Françoise Gouzi
Edward J. Gray
Stefan Hynek
Dimitar Ilkov Iliev
Maria Ilvanidou
Alba Irollo
Melina Leonie Jander
Adeline Joffres
Daniel Kurzawe
Michael Kurzmeier
Jakob Lenardič
Eliza Papaki
Vlad Popovici
Riccardo Antonio Celestino Pozzo
Marco Raciti
Nanette Rissler-Pipka
Amelia Sanz
Andrea Scharnhorst
Walter Scholger
Kristen Michelle Schuster
Eiríkur Smári Sigurðarson
Maria Spiliotopoulou
Alexander Steckel
Tomasz Umerle
Ubbo Veentjer
Lukas Weimer
Tanja Wissik
Tihomir Zivic