

Introduction: Breaking Silos, Connecting Data: Advancing Integration and Collaboration in Digital Humanities

Estelle Bunout¹, Tom Gheldof², and Thomas Smits³

¹University of Luxembourg

²KU Leuven

³University of Amsterdam

The 11th DH Benelux Conference, held at the Irish College in Leuven from 4 to 7 June 2024, gathered Digital Humanities scholars, developers, librarians, and cultural heritage professionals under the theme **Breaking Silos, Connecting Data: Advancing Integration and Collaboration in Digital Humanities**. This theme captured a prominent concern of our field today: the need to transcend disciplinary and institutional boundaries, while addressing the fragmentation of digital resources, infrastructures, and communities. The conference brought together a record number of contributions – 151 submissions, of which 111 were accepted - that examined how data, tools, and methods can be connected across space, time, and domain: from ancient prosopography to 20th-century propaganda archives, and from text analysis to computer vision. The theme was also particularly fitting for our venue, where the historical, archival and academic landscape of Leuven offered a natural setting for a meeting focused on linking data and scholars. The pre-conference program reflected the theme with workshops on collaborative tools such as Transkribus, the CLARIAH-VL tool suite, and sessions about making scholarly collections of letters accessible for a wide audience, experience and challenges with named entities, and opening up born-digital data for researchers. During the conference, the two keynote speakers also framed the conference’s theme by highlighting both the present reality and possible futures of the digital humanities. In her opening keynote *Medievalist in the Loop: Putting the Humanities back into the Digital Humanities*, Katarzyna Kapitan argued that, in an increasingly digital research landscape, virtually all humanities scholars already engage in digital scholarship to some extent. Rather than posing a threat to traditional methodologies, she emphasized that digital and conventional approaches function best in tandem: macro-scale analyses enabled by digital tools gain interpretive depth through traditional, close-reading perspectives, while established scholarly methods benefit from the new insights and efficiencies offered by computational techniques. The closing keynote by Manuel Burghardt (*The Future of Digital Humanities – Beyond Humanities?*) looked ahead and proposed a provocative vision for the field. He suggested that many branches of DH may gradually reintegrate into their respective humanities domains as digital methods become fully normalized, while Computational Humanities - with

its explicit modelling practices and formal, quantitative methods - may continue to develop as a distinct discipline. This forward-looking perspective underscored the characteristic openness of the field, as well as its willingness to question and redefine its own boundaries in response to evolving research practices. Across the three conference days, presentations, demos, and posters further explored how interoperability can foster more sustainable and inclusive digital scholarship. The papers published in this special issue represent a cross-section of that shared effort, demonstrating the diversity of approaches through which the DH Benelux community continues to advance the collaborative nature of our field.

Several contributions of this special issue turn their attention to the theme of data integration and network analysis, fields that lie at the core of research in digital history. These papers demonstrate how linking historical data can uncover new perspectives on social structures, migration, and intellectual networks. In their paper *Modelling Jewish Migration in Antiquity: connecting archaeological, inscriptional, and historical data from around the Mediterranean*, Stefan Dingemans, Tijmen C. Baarda, Berit Janssen, Arjan Mossel, and Leonard V. Rutgers reconstruct networks of Jewish communities across the Mediterranean from the 2nd century BCE to 600 CE. By integrating archaeological, epigraphic, papyrological, and literary datasets, the authors show how Jewish presence and movement can be visualised and quantified. Equally ambitious in scope, *STUDIUM.AI: datafying and connecting the 'webs of knowledge' around the premodern University of Leuven (1425–1797)* by Yann Ryan, Margherita Fantoli, Yanne Broux, and Violet Soen tackles the challenge of connecting several databases on the Old University of Leuven. Through standardization of person names, bibliographic records, and georeferenced locations, this project constructs a relational database infrastructure that dissolves the institutional silos of Leuven University. The authors demonstrate how linking these databases can provide a new perspective on early modern academic culture.

Other contributions examine the potential of automated text analysis, entity extraction, and AI-assisted annotation, showing how computational pipelines can structure vast and fragmented corpora. In *Accessing the Republic. Entity extraction from the resolutions of the Dutch States General*, Marijn Koolen, Esger Renkema, Nienke Groskamp, Frank Smit, Jirsi Reinders, Ronald Sluijter, Rik Hoekstra, and Joris Oddens describe how they developed a multi-layered pipeline that applies named-entity recognition to nearly one million early modern resolutions. Their approach enables the identification of individuals, committees, and dates in a body of texts that for centuries resisted systematic analysis, thus opening new avenues for the study of governance, deliberation, and political participation in the Dutch Republic. Maria Dermentzi, Mike Bryant, Fabio Rovigo, and Herminio García-González pursue a different yet complementary goal with their paper on *Multilingual Automated Subject Indexing: A Comparative Study of LLMs vs Alternative Approaches in the Context of the EHRI Project*, where they evaluate how large language models can assist archival indexing across languages. By comparing traditional lexical and statistical methods with fine-tuned and zero-shot LLM-based approaches, the authors highlight not only the accuracy of AI-driven classification but also its implications for cross-lingual discovery and the crucial role of human validation in curating Holocaust-related collections. Similarly, Mari Wigham, Rana Klein, Marjet Broelsma, and Roeland Ordelman (*Using AI to put together the wartime propaganda puzzle*) combine OCR, ASR, and NER technologies to link together radio broadcasts, their transcriptions, and contemporary newspaper materials on Dutch World War II propaganda. Despite challenges related to tran-

scription quality, their study demonstrates the integrative power of AI in restoring fragmented historical sources to analytical coherence. Florentina Armaselu's *Small-Scale Testing on Generative AI and Post-OCR Correction in Historical Datasets* offers a more experimental reflection, testing generative AI models for post-OCR correction on 18th-century French monographs. Her findings reveal how prompt engineering can balance historical fidelity with machine-readability, highlighting both the promise and risks, including the necessary ethical caution, when applying generative AI for text restoration.

Several papers discussed the infrastructural dimension of digital scholarship. In *Linking Person Observations FAIRly: PiCo in Practice*, Sytze H. J. Van Herck, Ivo Zandhuis, Richard Zijdemans, and Rick J. Mourits, present the Persons in Context (PiCo) ontology. They explain how a FAIR-compliant model can connect individual-level historical observations across datasets, creating provenance links that strengthen collaboration between archival and research institutions. Bert Aernouts, Mehmet Celik, and Joris Colla document the renewal of ODIS in *Breaking Down Barriers: The Transition of ODIS from a Relational to a Triple Store Database*. They not only detail their shift from a relational to a triple store architecture, but also explain how new SHACL-based workflows improve interoperability and enable richer forms of network visualisation and GIS integration. In a related exploration of audiovisual data (*Linking Het Amsterdams Stadsjournaal: A Case Study in Emerging Linked Open Data (LOD) Approaches to Audio-Visual Heritage*), Meg Weijers and Christian Olesen address the conceptual and technical complexities of connecting dispersed film materials across institutions through Wikibase/Wikidata and the CLARIAH Media Suite. Their study identifies the benefits of collaborative data modelling but also interrogates the interpretive tensions that accompany open linking in the heritage domain.

Equally significant are contributions that interrogate visualisation and bias in cultural heritage collections. *Herbaria Heritage: Visualising Colonial Bias in Natural History Collections* is the telling title of the paper in which Sakura Morales Furuta, Ana Reviejo Salamanca, Michaela Todd, Lise Stork, and Andreas Weber use interactive PowerBI dashboards to expose colonial extraction patterns in five million plant specimens. Their paper challenges the neutrality of data visualisation and re-situate natural history collections within their imperial context. Julie M. Birkholz, Benoît Crucifix, Erwin Dejasse, Sébastien Hermans, Krishna K. Thirukokaranam Chandrasekar, and Bas Vercruyssen show how computer vision (YOLOv8) can automatically detect and classify visual narratives in interwar Belgian periodicals (*Uncovering Interwar Comics: The Challenge of Labeling Graphics in Belgian Weekly Magazines*). Their paper provides a window into the visual and political culture of the 1930s and illustrates how visual data can be analyzed at scale. Katja Tereshko, Marijn Koolen, and Eva Viviani combine stylometric and reader-response methods in their paper *Reader response on the unconventional narrative tense in different genres*. They analyzed over 18,000 Dutch-language novels: this demonstrates how narrative experimentation with tense shapes readers' emotional and cognitive engagement. Julie Verlinden, Thorsten Mahieu, and Toon Goedemé's *Can stones be made (artificially) intelligent?* tackle similar methodological questions with an example from archaeology: the authors apply computer vision to detect recurring architectural motifs in Roman Asia Minor, combining pattern recognition with archaeological interpretation to reveal design logics invisible to the human eye.

Together, these 13 contributions exemplify the conference's overarching concern: how to overcome the disciplinary, technical, and institutional silos that continue to fragment Digital Humanities research. The authors address the challenges of working

with incomplete, fragmented, and heterogeneous datasets dispersed across institutions and formats. Their papers reveal how the analysis of data depends not only on technical interoperability but also on conceptual and ethical awareness.

Many contributions comply with FAIR and Open Research principles, showing how transparent and reusable infrastructure can underpin sustainable scholarship. Connecting data is a form of interpretation: it reshapes how researchers perceive relationships within and across collections, revealing patterns and biases that were previously hidden. Breaking silos, in this sense, is not only a technical challenge that can be tackled with a diversity of methods but also an interpretive one, which requires new conceptual frameworks.

At the end of this introduction, we would like to acknowledge and thank the local organising committee (Leah Budke, Demmy Verbeke, Rebecca Burke, Margherita Fantoli, Tom Gheldof and all the student volunteers) who made DHBenelux 2024 a temporary home for its participants. We also like to thank the local chairs and the reviewers, which made selecting the presentations and the papers for this volume an easy and enjoyable task.