Umanistica Digitale – ISSN: 2532-8816 – n. 22, 2026: Special Issue
G.M. Di Nunzio, M. Cornia – *Introduction to the Special Issue. From Data to Heritage: Semantic Enrichment, Archives, and Digital Libraries*
DOI: https://doi.org/10.60923/issn.2532-8816/24236

Introduction to the Special Issue. From Data to Heritage: Semantic Enrichment, Archives, and Digital Libraries

Giorgio Maria Di Nunzio

Department of Information Engineering, University of Padova, Padova, Italy
giorgiomaria.dinunzio@unipd.it

Marcella Cornia

Department of Education and Humanities, University of Modena and Reggio Emilia, Reggio Emilia, Italy
marcella.cornia@unimore.it

Abstract

This special issue of Umanistica Digitale, developed in conjunction with IRCDL 2025 (Information and Research science Connecting to Digital and Library science), presents extended versions of selected contributions from a conference explicitly designed to connect two complementary tracks: Computer Science Foundations for Digital Libraries and Digital Humanities. The issue showcases current research at the intersection of semantic technologies, digitization workflows, and computational analysis of cultural data. These contributions articulate a shared agenda for strengthening collaboration between digital-library research and humanities scholarship, emphasizing openness, accessibility, methodological transparency, and reusable infrastructures.

Keywords: IRCDL 2025, Digital Libraries, Semantic Web, Digital Archives

Questo numero speciale di Umanistica Digitale, realizzato in collaborazione con il convegno internazionale IRCDL 2025 (Information and Research science Connecting to Digital and Library science), presenta versioni estese di contributi selezionati provenienti da un evento costruito esplicitamente per mettere in dialogo due filoni complementari: "Computer Science Foundations for Digital Libraries" e "Digital Humanities". Il numero mette in evidenza le ricerche più recenti all'intersezione tra tecnologie semantiche, flussi di lavoro di digitalizzazione e analisi computazionale dei dati culturali. I contributi delineano un'agenda condivisa volta a rafforzare la collaborazione tra la ricerca sulle biblioteche digitali e gli studi umanistici, ponendo l'accento su apertura, accessibilità, trasparenza metodologica e infrastrutture riusabili.

Parole chiave: IRCDL 2025, Biblioteche digitali, Web semantico, Archivi digitali

Introduction

This special issue of Umanistica Digitale grows out of the collaboration with IRCDL 2025 and collects selected, extended versions of contributions presented at the conference. IRCDL[1] (Information and Research science Connecting to Digital and Library science), established in 2005 (formerly Italian Research Conference on Digital Libraries), is an annual venue that brings together a broad community around Digital Libraries and related themes, spanning technical, practical, and social perspectives and welcoming both national and international participation. It explicitly positions Digital Libraries within a multidisciplinary ecosystem, computer science and information science as well as librarianship, archival and museum practices, social sciences, cultural heritage and the humanities, while maintaining links to major international initiatives in the field and to adjacent communities (including AI for cultural heritage).

The 21st edition, IRCDL 2025 [1], was held at the Biblioteca Rizzi, University of Udine, on 20–21 February 2025, and was structured around two complementary tracks: 1) Computer Science Foundations for Digital Libraries, focused on core CS concepts for DLs, algorithms and AI methods for information retrieval and data management, architectures for large-scale collections, and applications in research and cultural-heritage preservation; and 2) Digital Humanities, devoted to the intersection of digital technologies and humanities research, including computational approaches to analyzing and preserving cultural artefacts, text-mining for large-scale literary analysis, and digital platforms for collaborative scholarship. The conference accepted regular papers, short papers, and extended abstracts, all reviewed by two or three program committee members, reflecting IRCDL's long-standing aim of connecting research communities across methods, objects, and professional contexts.

Building on this format and on the shared conviction that today's digital-library challenges increasingly sit at the crossroads of AI, archives, and humanities inquiry, Umanistica Digitale launched a dedicated call to strengthen and institutionalize this interdisciplinary dialogue in an editorial venue. The present special issue thus opens a new series/joint venture: a curated space where contributions born in IRCDL's two-track structure can be expanded, contextualized, and brought into closer conversation for the UD readership, with the broader goal of consolidating collaboration between computer science foundations for digital libraries and the questions, sources, and interpretive practices of Digital Humanities.

Overview of the Contributions

Across the seven contributions, the special issue shows how digital-library research and Digital Humanities can mutually reinforce each other: from semantic modelling and knowledge graphs that improve discovery and reuse, to datasets and pipelines that strengthen digitization and metadata quality, and to NLP methods that open new paths for scholarly analysis.

[7] investigate how a data-storytelling platform can support students and DH scholars in developing data and visualization literacy on Linked Open Data. By analyzing projects and user feedback, the study highlights a preference for linear narratives and "easy-to-consume" visualizations, with benefits for critical thinking and iterative learning, while also noting limitations due to incomplete data sources.

[1] https://ircdl.dei.unipd.it/

[3] address openness and accessibility in digital music archives, where rare and fragile materials demand solutions that enhance usability and inclusion. The paper discusses multilayer formats, especially IEEE 1599, and, through two cases from the Ricordi Historical Archive, shows how synchronizing notation, audio, media, and metadata can enable multimodal access as well as assistive and personalized experiences.

[5] start from the observation that existing semantic models often fail to capture the specificity of DH "products", reducing visibility and discoverability. The paper presents early versions of an ontology and knowledge graph designed to represent outputs such as digital editions, textual collections, linked open datasets, ontologies, and software, with the goal of improving findability, interoperability, and reuse.

[8] introduce a public dataset of Arabic-script title pages aimed at improving OCR, document-structure understanding, and metadata accuracy in digital libraries. The workflow combines a vision-language model to classify title pages with Google Vision AI OCR to produce ground truth later revised by experts, alongside plans to experiment with open-source alternatives.

[4] frame cultural-heritage access as an economic driver (e.g., tourism and services) and presents CHiPS&BITS as a knowledge-based approach to overcome limitations of current web solutions. With a focus on the History of Computing, it proposes dedicated ontologies and AI algorithms to deliver more advanced, user-centered, and personalized experiences.

[2] report an interdisciplinary effort to automatically annotate legal references in the Ordinary Gloss to the "Liber Extra", producing a navigable index that supports exploration and analysis tools. Beyond the specific case study, the paper emphasizes a methodological takeaway: meaningful, robust results can be achieved by small teams with limited resources, and it makes both data and code openly available.

[6] propose a methodology for intertextuality in patristic Latin that combines philological rigor with NLP. It integrates token-based classification with semantic annotation and retrieval via Latin BERT models fine-tuned using contrastive learning and hard negatives; on Augustine's "De Genesi ad litteram" and Latin biblical texts (Vulgate and pre-Vulgate), it reports improvements over baselines and outlines a scalable path to connecting explicit quotations and implicit allusions.

References

[1] Cornia, Marcella, Giorgio Maria Di Nunzio, Donatella Firmani, et al., eds. 2025. "Preface Summary." In *Proceedings of the 21st Conference on Information and Research Science Connecting to Digital and Library Science*, vol. 3937. CEUR Workshop Proceedings. CEUR. https://ceur-ws.org/Vol-3937/#preface.

[2] Esuli, Andrea, Vincenzo Roberto Imperia, and Giovanni Puccetti. 2026. "Automatic Annotation of Legal References (Allegationes) in the Liber Extra's Ordinary Gloss." *Umanistica Digitale*, no. 22 (January): 139–56. https://doi.org/10.60923/issn.2532-8816/22163.

[3] Faschi, Vanessa, Federico Avanzini, and Luca Andrea Ludovico. 2026. "Fostering Openness and Accessibility of Music Archives through Multilayer Formats." *Umanistica Digitale*, no. 22 (January): 25–43. https://doi.org/10.60923/issn.2532-8816/21759.

[4] Ferilli, Stefano, Eleonora Bernasconi, Domenico Redavid, and Giorgio Maria Di Nunzio. 2026. "Knowledge Graphs for the Web Economy: The CHiPS&BITS Project on Cultural Heritage." *Umanistica Digitale*, no. 22 (January): 113–38. https://doi.org/10.60923/issn.2532-8816/22159.

[5] Giacomini, Sebastiano, Chiara Martignano, Giorgia Rubin, et al. 2026. "ATLAS: A Knowledge Graph to Enhance the Findability of International Scholarly Research on the Italian Digital Cultural Heritage." *Umanistica Digitale*, no. 22 (January): 45–76. https://doi.org/10.60923/issn.2532-8816/21765.

[6] Mambelli, Anna, Laura Bigoni, Davide Dainese, et al. 2026. "The Biblical Heritage in Ancient Latin Christian Literature: Advancing Intertextual Mapping Through Sentence Embeddings." *Umanistica Digitale*, no. 22 (February): 157–86. https://doi.org/10.60923/issn.2532-8816/22160.

[7] Renda, Giulia, and Marilena Daquino. 2026. "Do Information Visualisation and Semantic Web Get on Well? A Study on the Usage of Storytelling Platforms to Improve Data Literacy Skills." *Umanistica Digitale*, no. 22 (January): 1–23. https://doi.org/10.60923/issn.2532-8816/21755.

[8] Vigliermo, Riccardo Amerigo, Giovanni Sullutrone, Sonia Bergamaschi, and Luca Sala. 2026. "The 'Digital Maktaba LP': Proposing a Comprehensive Dataset for Arabic Script OCR Title Pages in the Context of Digital Libraries and Religious Archives." *Umanistica Digitale*, no. 22 (January): 77–111. https://doi.org/10.60923/issn.2532-8816/22019.