Secondary Publication



Gradl, Tobias; Fruth, Leon; Geißler, Nils; u. a.

Future Applications : The Text+Registry as Reusable Technical Component Beyond Resource Cataloguing

Date of secondary publication: 28.10.2026

Accepted Manuscript (Postprint), Conferenceobject

Persistent identifier: urn:nbn:de:bvb:473-irb-117723x

Primary publication

Gradl, Tobias; Fruth, Leon; Geißler, Nils; u. a. (2026): Future Applications: The Text+Registry as Reusable Technical Component Beyond Resource Cataloguing, in: Aleksandra Rykowska (Hrsg.), European Association for Digital Humanities 2026: "Linking Europe: Digital Humanities Without Borders“, Zenodo, S. 166–174, doi: 10.5281/zenodo.23363983

Legal Notice

This work is protected by copyright and/or the indication of a licence. You are free to use this work in any way permitted by the copyright and/or the licence that applies to your usage. For other uses, you must obtain permission from the rights-holders.

This document is made available under a Creative Commons license.



The license information is available online:
https://creativecommons.org/licenses/by/4.0/legalcode

eadh
EUROPEAN ASSOCIATION FOR DIGITAL HUMANITIES

Future Applications — The Text+ Registry as Reusable Technical Component Beyond Resource Cataloguing

Tobias Gradl [1], Leon Fruth [1], Nils Geißler [2], Kilian Hensen [2], and Daniela Schulz [3]

[1]Media Informatics Group, University of Bamberg, Germany · [2]Cologne Center for eHumanities (CCeH), University of Cologne, Germany · [3]Herzog August Bibliothek Wolfenbüttel, Germany

1 Introduction

The design of digital humanities' research software, tools, or infrastructures usually follows domain-specific requirements, resulting in solutions that address particular scholarly needs but often lack transferability across contexts. The Text+ Registry, developed within the Text+ consortium of the German National Research Data Infrastructure (NFDI), offers an alternative approach. Its technical architecture is engineered not only to address immediate cataloguing challenges related to editions, lexical resources and collections, but also to serve as a reusable infrastructural component applicable to fundamentally diverse scholarly and institutional contexts. This paper examines how the Registry's three architectural principles—metamodelling, model-driven design and information layering—create possibilities for application scenarios that extend beyond its original purpose. It offers a model for the sustainable and adaptable development of infrastructure components—not only in the digital humanities.

2 Architectural Foundations Implications

The Registry's metamodelling approach (Gradl, Fruth, and Henrich 2025) allows multiple, structurally distinct data models to be handled simultaneously, thus eliminating the need for architectural redesign. Unlike traditional database systems, which implement a fixed schema, the Registry operates at a meta-level, where the data models themselves become configurable components. These models—in the case of Text+ currently encompassing textual resources in the form of editions, collections, and lexical resources—are formalised in YAML specifications, which are stored in a version-controlled repository. These specifications are then transformed into application-specific representations through a generic metamodelling language. This abstraction layer means that adding new resource types, modifying existing schemas or adapting

to evolving community standards requires merely the updating of model definitions rather than a revision of application code.

Thus, the implications of this design principle extend beyond technical elegance. In practice, metamodelling addresses a fundamental tension in the realm of digital humanities infrastructure: the need to accommodate disciplinary specificity while enabling cross-domain interoperability. The editions data model (Fruth et al. 2025; Gradl et al. 2024), developed through extensive consultation with domain experts and incorporating standards from library science and editorial theory, differs substantially from parallelly implemented models for lexical resources, dictionaries and linguistic databases. Yet both coexist within the same technical infrastructure, sharing common entity types while maintaining domain-specific granularity. Cross-domain queries operate through DataCite mappings that identify semantic equivalences across models, while domain-specific metadata remains fully preserved and searchable within its native structure.

Model-driven design complements meta-modelling by generating software components directly from formal models rather than through manual coding. This approach, borrowed from software engineering domains requiring strict formalization and technology independence, proves particularly valuable for humanities infrastructure where sustainability demands outlast specific technological stacks.

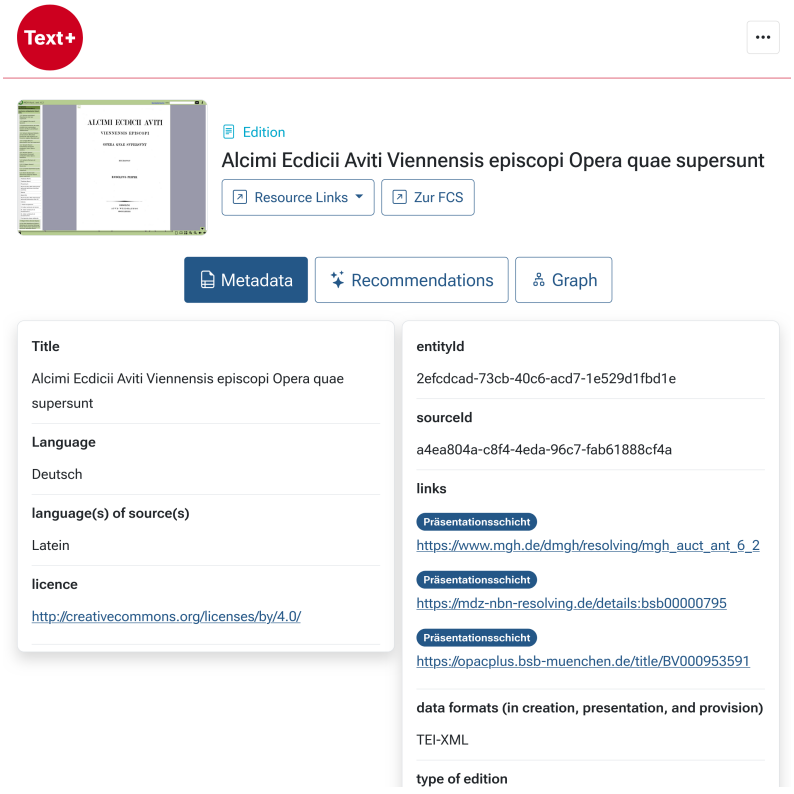


Figure 1: An openMGH edition in the Registry search interface

When the Registry ingests metadata from heterogeneous sources like research information systems or various catalogues and databases, the transformation pipelines are generated from mappings between source and target models defined in the Data Modelling Environment (DME) (Gradl 2024). Changes to either source formats or target schemas merely require updating these mappings rather than rewriting transformation logic, dramatically reducing maintenance burden and enabling the system to evolve alongside changing standards. As an example, consider the extension of the Registry's editions landscape by openMGH, which provides open-access TEI-XML downloads of Monumenta Germaniae Historica critical editions covering the Middle Ages in Germany and Europe. Based on the listing of MGH editions available in openMGH, data modelling in the DME captures the full pipeline: parsing the HTML listing, extracting TEI-XML, and following referenced RDF and IIIF descriptions to compile edition descriptions that can be imported and searched in the Text+ Registry. Figure 1 shows an openMGH edition in the Registry's search interface, illustrating how structured metadata is surfaced and made explorable.

Information layering constitutes the Registry's third foundational principle, addressing metadata heterogeneity not through normalization but through compositional synthesis. When multiple sources provide information about the same resource, the Registry preserves each as a distinct, attributed layer rather than merging them into a single, decontextualized record. A predefined hierarchy determines which layer takes precedence for display purposes, however, all original metadata remains accessible with full provenance information. Manual expert curation adds additional layers without overwriting automatically harvested data, enabling distributed enhancement while maintaining audit trails. This provenance-aware composition transforms metadata integration from a lossy flattening process into an additive knowledge-building activity where each contribution, whether automated aggregation or human curation, enriches the collective resource description. This design directly operationalises the FAIR principles of accessibility and reusability, ensuring that metadata provenance remains traceable and machine-actionable across all contributing layers (Wilkinson et al. 2016).

3 Beyond Cataloguing: Application Scenarios for Reusable Infrastructure

The Registry's architectural principles enable application scenarios that transcend its original cataloguing purpose. The fundamental challenge these principles address—managing heterogeneous, evolving data structures while maintaining provenance and enabling both automated processing and human curation—recurs across diverse contexts in digital humanities and beyond.

3.1 Scenario A: MIDA Registry

The long-term project 'Modern India in German Archives' (MIDA) catalogues German archival material on the history of modern India (1706–1989/90) and German-Indian relations. The aim is to systematically record relevant holdings, relate them and make them permanently available to the international research community. In this context, sources previously considered marginal from a national or colonial his-

torical perspective and scattered among various holdings, are now in the centre of investigation.

With the end of the active funding phase, the MIDA project requires a solution enabling a transition to a flexible and openly accessible cataloguing infrastructure, in order to continue providing access to project outcomes. For this purpose, a dedicated Registry instance was deployed, with data models defined to exactly reflect MIDA's core semantic units—archive, holding, and archival material. The three architectural principles of the Registry facilitated the migration from the project's existing database without any information loss. The MIDA Registry is currently being evaluated (see Figure 2), and will be made publicly accessible upon completion of this process.

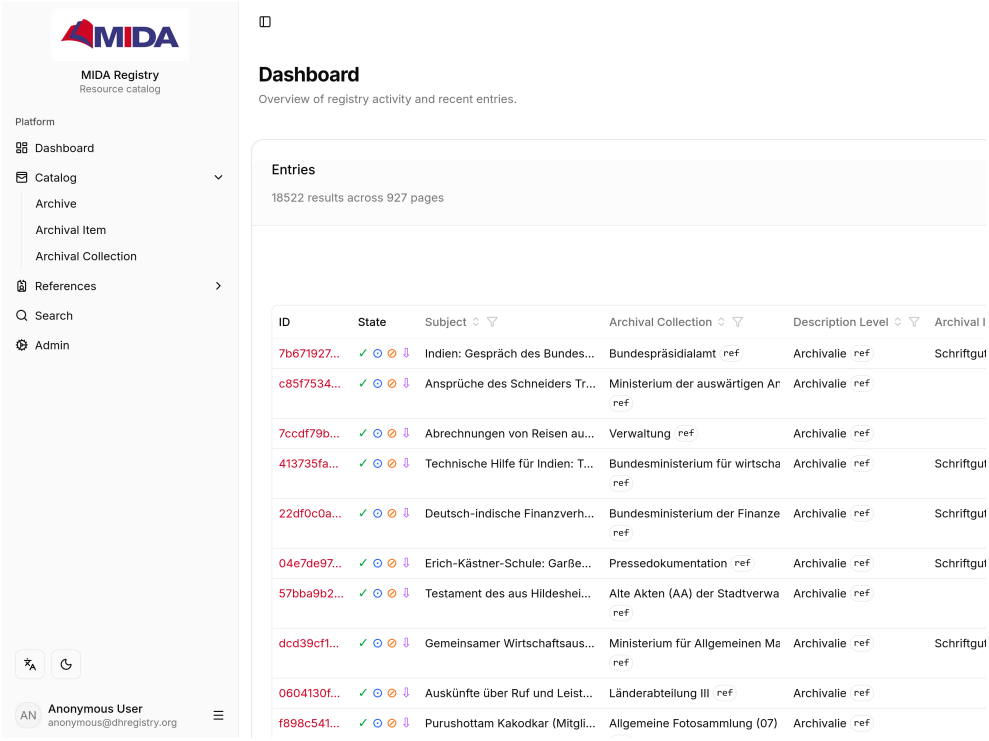


Figure 2: Archival material in the administrative interface of the MIDA Registry

3.2 Scenario B: The Registry as Catalogue for Research Data Management (RDM) Infrastructures in Germany as Part of the National Research Data Infrastructure (NFDI)

Various platforms and indexes have been created within the Research Data Management (RDM) community to help researchers find support. These include the NFDI helpdesk overview, platforms such as forschungsdaten.info and forschungsdaten.org, the Social Sciences & Humanities Open Marketplace (SSHOMP) (SSH Open Marketplace 2026), and the NFDI-N(HAW)igator (Eilert and Reiter 2024) or come2data (Come2Data 2026). These successful efforts, however, result in platforms facing lim-

itations in either scope or functionality. Specifically, essential features such as advanced search and filtering, interoperability with external systems, user-centric service descriptions, and machine learning-powered recommendations are currently underdeveloped.

As a reaction, an additional branch of the Registry will be used to establish a Helpdesk Catalogue within the NFDI. This catalogue will integrate RDM stakeholder support from across the entire NFDI as a RDM network (Engel et al. 2024). Its metadata schema will be based on the DataCite Resource Type "Service," augmented by community-requested fields identified through surveys, and informed by established RDM service metadata models (e.g., Lemaire et al. (2025)). Furthermore, controlled vocabularies will be incorporated.

4 Technical Implementation and Interoperability Patterns

The Registry's technical architecture establishes patterns for interoperability that avoid common pitfalls of integration projects. Rather than requiring participants to adopt uniform metadata standards before integration, the system accepts heterogeneous inputs and performs alignment post-hoc through configurable mappings. Rather than centralizing data storage and requiring migration from existing systems, it operates as a metadata aggregator that harvests from distributed sources via standard protocols (OAI-PMH, RESTful APIs, SPARQL endpoints) while leaving authoritative data in place. Rather than imposing top-down schemas, it enables community-driven model development where domain experts formalize their own conceptual frameworks in collaboration with technical implementers.

5 Facilitating Pluralism

The Text+ Registry's significance extends beyond its immediate functionality in cataloguing scholarly resources to its demonstration of an architectural philosophy for digital humanities infrastructure. Rather than seeking universal solutions that compress disciplinary diversity into uniform structures, it enables productive pluralism where multiple conceptual frameworks coexist, interoperate, and inform each other without requiring convergence. The metamodelling approach acknowledges that different communities legitimately conceptualize their materials according to different logics following an approach according to which technical infrastructure should accommodate these differences rather than forcing premature standardization.

This approach has implications for ongoing debates about digital humanities infrastructure, particularly tensions between local customization and global interoperability, between disciplinary specificity and cross-domain integration, between expert curation and automated processing. The Registry suggests these need not be zero-sum tradeoffs but can be held in productive tension through architectures that support multiple levels of abstraction, preserve provenance through all transformations, and enable participation at varying levels of technical expertise and institutional support.

As digital humanities infrastructures mature, the question shifts from whether to build shared systems to how to build them in ways that serve diverse communities equitably, adapting to changing scholarly needs and technological possibilities, and sustaining themselves through active engagement rather than institutional inertia.

The Registry provides one solution, grounded in specific requirements of text- and language-based research in the context of Text+ but with architectural principles potentially applicable across humanities computing. The paper will focus on actual as well as potential applications of the Registry based on its architectural principles, and thus contribute to understanding how to build infrastructure that connects without homogenizing, that scales without losing specificity, and that serves scholarship rather than constraining it.

Acknowledgements

This document was created in the context of the work of the association German National Research Data Infrastructure (NFDI) e.V. NFDI is financed by the Federal Republic of Germany and the 16 federal states, and the consortium Text+ is funded by the DFG - project number 460033370. The authors would like to thank for the funding and support. Furthermore, thanks also include all institutions and actors who are committed to the association and its goals.

References

Come2Data (July 2026). *Come2Data Datenkompetenzkarte Prototyp*. URL: https://umap.openstreetmap.de/en/map/come2data-datenkompetenzkarte-prototyp_79991.

Eilert, Eva and Martin Reiter (2024). *NFDI-N(HAW)igator*. DOI: 10.5281/ZENODO.13973397.

Engel, Judith, Patrick Helling, Robert Herrenbrück, et al. (2024). *Working Group Charter. RDM Helpdesk Network*. en. DOI: 10.5281/ZENODO.14035822.

Fruth, Leon, Nils Geißler, Daniela Schulz, and Tobias Gradl (2025). "Cataloguing Editions and Other Resources in One Unified System: The Case of the Text+ Registry". In: *Proceedings of the Digital Humanities Congress 2024*. Sheffield: Digital Humanities Institute. URL: https://www.dhi.ac.uk/books/dhc2024/fruth-et-al/.

Gradl, Tobias (2024). "Filling the gaps: Facilitating access to enriched research data". en. In: *DH2024 Book of Abstracts*. Ed. by Jajwalya Karajgikar, Andrew Janco, and Jessica Otis. Zenodo. DOI: 10.5281/ZENODO.13752750.

Gradl, Tobias, Leon Fruth, and Andreas Henrich (2025). "The Text+ Registry: Federating Research Data Catalogues for the Digital Humanities". In: *Linking Theory and Practice of Digital Libraries*. Springer Nature Switzerland, pp. 378–394. ISBN: 9783032054098. DOI: 10.1007/978-3-032-05409-8_22.

Gradl, Tobias, Christoph Kudella, Harald Lordick, and Daniela Schulz (2024). "Towards a Registry for Digital Resources – The Text+ Registry for Editions". In: *Datenbank-Spektrum* 24.2, pp. 151–160. ISSN: 1610-1995. DOI: 10.1007/s13222-024-00479-0.

Lemaire, Marina, Patrick Helling, Boris Jacob, Stefan Kellendonk, and Rouven Schabinger (2025). *Ein konsolidiertes Metadatenschema für FDM-Services*. DOI: 10.5281/ZENODO.15025880.

SSH Open Marketplace (July 2026). *CESSDA, CLARIN, & DARIAH-EU*. URL: https://marketplace.sshopencloud.eu/.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, et al. (Mar. 2016). "The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific Data* 3.1. ISSN: 2052-4463. DOI: 10.1038/sdata.2016.18.