

Submission for a poster at the **HMC Conference 2026 Topic: Metadata in Action**
in Heidelberg, from April 28 to 30, 2026 held by the Helmholtz Association at the German Cancer Research Center
(DKFZ), Track 2: Empowering Research Communities: Turning Metadata into Action

Navigating the Text+ Data Space: Metadata & the Search Experience

Thomas Eckart [1], Stefan Buddenbohm [2], Leon Fruth [3],
Tobias Gradl [3], Maximilian Hebeis [3], Uwe Kretschmer [1], Timm Lehmborg [4]¹

[1] Saxon Academy of Sciences and Humanities in Leipzig, [2] Göttingen State and University Library,
[3] University of Bamberg, Chair of Media Informatics, [4] Academy of Sciences and Humanities in Hamburg

Research infrastructures serve as vital interfaces between data providers and user groups. Within the German National Research Data Infrastructure (NFDI), the Text+ consortium focuses on text- and language-based research data across three core domains: collections, scholarly editions, and lexical resources. Since 2021, Text+ has been developing a federated data space that integrates geographically and organizationally distributed data centers to ensure the FAIR (Findable, Accessible, Interoperable, Reusable) and transparent provision of resources.

A central pillar of the Text+ Data Space is the consistent application of metadata, including controlled vocabularies and authority files. These elements are essential for enhancing resource discovery and findability. By connecting an extensive resource inventory based on shared vocabulary and referenced authority file entities the researchers are able to navigate a complex digital landscape. The use of authority files and persistent identifiers (PIDs) offers additional advantages: they provide stable references to disambiguated entities and ensure interoperability within a heterogeneous and evolving environment.

However, establishing such a large-scale infrastructure involves significant challenges. These range from the structural heterogeneity of data formats to the semantic diversity of metadata—which must balance global applicability with project-specific requirements. Furthermore, a primary goal is the user-friendly design of services that cater to a wide range of researchers with varying levels of technical expertise.

This poster presents the search experience within the Text+ Data Space. Through practical examples, we demonstrate how researchers can effectively navigate the infrastructure and how the integration of metadata and authority files facilitates the discovery of text- and language-based research data.

¹ ORCID IDs: Thomas Eckart: 0000-0001-6692-0713; Stefan Buddenbohm: 0000-0002-3469-6101; Leon Fruth: 0009-0001-2128-3025 Tobias Gradl: 0000-0002-1392-2464; Maximilian Hebeis: 0009-0008-2531-3616; Uwe Kretschmer: 0000-0002-3685-6519, Timm Lehmborg: 0009-0001-3133-6282

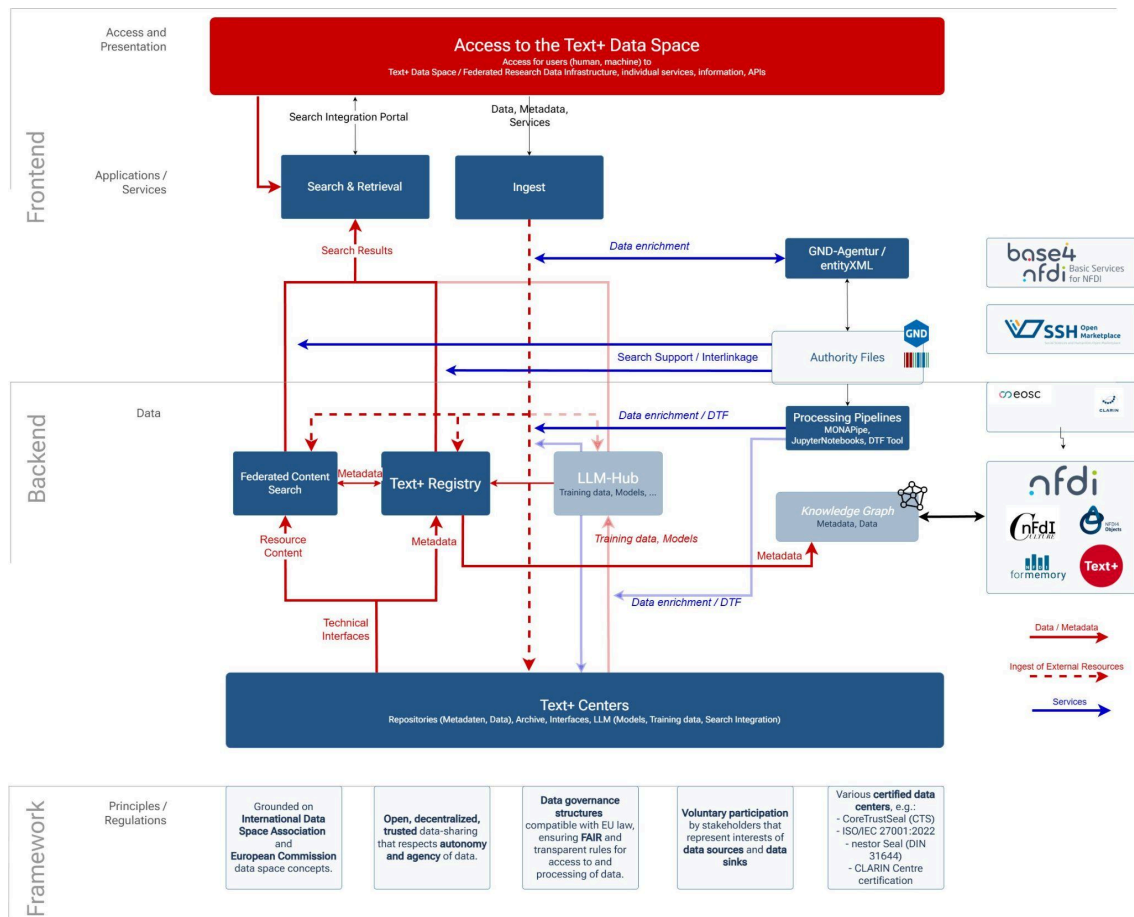


Figure 1: The Text+ Data Space as a slight variation of the Text+ Architecture². Grayed-out sections are on the roadmap and not implemented yet as of December 2025.

References

Data Spaces Support Centre (DSSC) (2025): Mission Statement. <https://dssc.eu/space/Partners/175472674>.

Fruth, L., Gradl, T., Hebeis, M. and Henrich, A. (2025): A Flexible Search System for Integrated Authority Data—ADISS. *Datenbank Spektrum* (2025). <https://doi.org/10.1007/s13222-025-00515-7>.

Glombiewski, N. et al. (2024): From Theory to Practice: Demonstrators of FAIR Data Spaces Across Different Sectors. <https://arxiv.org/abs/2412.04969>.

Gradl, T., Fruth, L. and Henrich, A. (2025): The Text+ Registry: Federating Research Data Catalogues for the Digital Humanities. In: Balke, WT., Golub, K., Manolopoulos, Y., Stefanidis, K., Zhang, Z. (eds) *Linking Theory and Practice of Digital Libraries*. TPDL 2025. Lecture Notes in Computer Science, vol 16097. Springer, Cham. https://doi.org/10.1007/978-3-032-05409-8_22.

Körner, E., Eckart, T., Helfer, F. and Kretschmer, U. (2025): An Enhanced Federated Content Search Infrastructure for the Humanities. In: *Selected papers from the CLARIN Annual Conference 2024*. 2025. <https://doi.org/10.3384/ecp216.05>.

Körner, E. and Eckart, T. (2025): Accessing linguistic content in distributed research environments. In: *Harmonizing language data: Standards for linguistic resources*, edited by Piotr Bański, Ulrich Heid and Laura Herzberg, De Gruyter, 2025, pp. 377-399. <https://doi.org/10.1515/9783112208212-015>.

² <https://text-plus.org/en/ueber-uns/architektur/>