

HUMMA.LV: Towards a Collaborative Digital Platform for Humanities and Arts in Latvia

Sandis LAIME¹, Sanita REINSONE²

¹ Institute of Literature, Folklore and Art of the University of Latvia, Mūkusalas 3, Rīga, Latvia

² University of Latvia, Faculty of Humanities, Raiņa blvd. 19, Rīga, Latvia

`sandis.laime@lulfmi.lv, sanita.reinsone@lu.lv`

ORCID 0000-0002-6529-6303, ORCID 0000-0003-1980-5450

Abstract. This overview paper discusses the HUMMA digital platform, detailing its evolution and broad objectives. Originally stemming from the Archives of Latvian Folklore, the platform has expanded to include diverse datasets from various institutions, enhancing accessibility and integration across the humanities and arts. HUMMA supports data storage, representation, and analysis, serving researchers, students, and the public interested in cultural processes and artistic expressions in Latvia. It also promotes citizen science and crowdsourcing to enrich its content, illustrating a commitment to collaborative, interdisciplinary data use in the humanities.

Keywords: digital infrastructure, open data platform, humanities data, digital humanities history, collaborative digital humanities.

1. Introduction

In recent decades, digital infrastructures have become indispensable to humanities research, enabling unprecedented access to and engagement with humanities and cultural heritage information. These infrastructures serve as essential tools for preserving, organizing, and analyzing diverse materials. By offering centralized, searchable, compatible, and open platforms and tools, they facilitate the systematic study and interpretation of complex information, supporting the development of new methodologies and advancing knowledge in the field.

Beyond their academic applications, humanities digital infrastructures also play a critical role in public engagement and societal impact. They empower cultural organizations, educational institutions, and communities to collaborate in the exploration, preservation and interpretation of cultural artifacts. Participatory digital tools, as well as citizen science and humanities crowdsourcing initiatives have expanded the reach of the humanities digital infrastructures, making them hubs for collaborative knowledge production and cultural dialogue. These developments highlight the intersection of humanities research with broader societal and technological trends, underlining the transformative potential of digital tools in supporting research, fostering cultural understanding, and addressing contemporary challenges in heritage preservation and knowledge dissemination (Terras, 2022; Edmonton and Morselli, 2020).

The development of digital infrastructures is a process with its own distinct history, shaped by specific components, administrative and financial contexts, developmental trends, and the influence of individuals who have contributed to or guided its trajectory. Over time, the sustainability of these infrastructures has emerged as an increasingly critical concern (Barats et al., 2020; Tucker, 2022). Ensuring their longevity involves addressing challenges related to technological obsolescence, funding continuity, and the evolving needs of users. As digital infrastructures continue to grow in complexity and importance, their sustainability is not only a technical or financial issue but also a matter of strategic foresight, requiring collaborative governance, adaptable frameworks, and active engagement from diverse stakeholders to maintain their relevance and utility in the long term.

This overview paper introduces HUMMA, an example of such an infrastructure, showcasing how it integrates diverse datasets and tools to address the needs of humanities research and cultural heritage preservation.

2. Overview of HUMMA

HUMMA is a digital platform for the humanities and arts, developed and maintained by the Institute of Literature, Folklore, and Art at the University of Latvia (ILFA). The platform integrates datasets from a range of contributors, including the University of Latvia's Faculty of Humanities and Institute of Philosophy and Sociology, Riga Technical University Liepāja, Rēzekne Academy of Technologies, the Latvian Theatre Workers' Union, and other partner organizations. HUMMA actively fosters collaboration with other institutions in the humanities and incorporates citizen science and crowdsourcing initiatives to create, interact and refine its content. The platform's overarching objective is to aggregate, preserve, and provide open access to diverse data types related to the humanities and arts, particularly focusing on the cultural practices of Latvians, the Latvian diaspora, and Latvia's ethnic minorities.

HUMMA's functionality encompasses data aggregation from various domains within the humanities and arts, including folklore, history, ethnomusicology, literary studies, and theater studies. Beyond serving as a repository, it offers tools for data representation as collections, map-based visualization, structured search, download of datasets, and analytical processing. Designed to serve a diverse audience, the platform supports researchers, students, and others in accessing structured and accessible information on topics related to culture and the arts.

3. History of HUMMA

The origins of HUMMA date back to the period between 2014 and 2016, when the Institute of Literature, Folklore, and Art initiated the development of a digital archive for the Archives of Latvian Folklore (ALF), garamantas.lv. This initiative aimed to consolidate and systematize ILFA's previously fragmented digital resources, enabling their institutional integration (Reinsone and Laime, 2022). During this foundational stage, a data structure was designed to facilitate the processing of folklore materials, and existing digitized archival materials, along with their associated metadata, were incorporated. Among these were the "Krātuve" database, which contained part of the

systematic index of ALF, as well as digitized photographs, drawings, and audio recordings. In 2015, the “Proverbs Database,” comprising the digitized ALF proverb card catalogue, was integrated into *garamantas.lv*. The project advanced further in early 2016 with the creation of the “Valodas talka” manuscript transcription platform, developed in collaboration with the Latvian National Commission for UNESCO. This initiative represented a milestone by introducing significant public involvement in the processing and enrichment of the digital archive’s content.

Concurrently, ILFA integrated biographical and bibliographical data from a literary database developed in previous years into *garamantas.lv*. This initiative also facilitated the creation of a complementary resource which has operated since 2017 under the name “Latvian Literature Digital Resource” at *literatura.lv* (Eglāja-Kristsons 2021; Eglāja-Kristsons and Graudiņa 2024). This expansion was instrumental in establishing a unified data structure for ILFA’s digital resources, addressing the challenges of organizing and integrating previously created digital resources. By consolidating existing repositories, it created a systematic framework that allowed ILFA to manage its humanities and cultural heritage data more effectively while enhancing the platform’s capacity to meet the diverse demands of humanities and arts research.

Between 2017 and 2020, the development of HUMMA prioritized the expansion of the platform with digitized ALF materials and the enhancement of its overall functionality. Significant advancements during this period included the introduction of a digital participation module, which enabled the recording and preservation of audio data within a dedicated collection created on the digital platform. Additionally, the ALF survey module, now the primary tool for enriching ALF’s collections with contemporary material, was implemented. These efforts culminated in the integration of advanced tools, such as a Places database, mapping tools, and visualization capabilities, which significantly expanded the platform’s analytical potential. A text annotation tool was also developed, enabling the generation of maps based not only on metadata but also on textual references to geographic locations. Furthermore, the platform incorporated an extensive open dataset from the Latvian Geospatial Information Agency’s Placename Database, supplemented by mapping data sourced from the State Land Service’s address register, Google Maps, and OpenStreetMap. These enhancements advanced the platform’s capacity to support complex research needs.

Prior to 2020, ILFA’s digital resources were hosted on separate platforms, such as *garamantas.lv* for folklore materials and *literatura.lv* for literary data. In 2020, the State Research Program “Digital Resources of the Humanities” launched the project “Digital Resources for Humanities: Integration and Cooperation,” which consolidated all ILFA digital resources into a unified platform, initially named *dati.lulfmi.lv*. This integrated resource provided centralized access to ILFA’s datasets, incorporating advanced search capabilities and enabling the download of text corpora for research purposes. Later that year, the platform was rebranded as HUMMA to reflect its interinstitutional scope and its emphasis on serving the broader humanities and arts community, while intentionally avoiding alignment with specific institutional identities.

In recent years, HUMMA has significantly advanced its efforts to integrate resources across platforms. For example, an API interface facilitated the incorporation of author data from *literatura.lv* into the Latvian National Library’s “Latvian Prose Counter” (<https://proza.lnb.lv>). Similarly, data from the privately developed <https://izrades.lv> database of theatrical performances were integrated into HUMMA, along with personal

records of theatre professionals, which were subsequently consolidated into the database and website of the Latvian Theatre Workers' Union at <https://ltds.lv>.

4. The infrastructure of HUMMA

The HUMMA infrastructure comprises several key components:

- A unified, non-public interface for data editing and administration, supporting multiple levels of user access rights based on the nature and sensitivity of the data.
- Sixteen interrelated databases designed according to a unified data model. These include Repositories and Collections databases for data structuring and specialized databases for research sources and their metadata, such as Files, Text Units, Audio and Video Units, Images, Literary Works, Performances, Dramaturgy, Awards, as well as databases for Persons, Organizations, Places, Individual Events, and Collective Events.
- A unified interface for data browsing and querying, available through the open data platform humma.lv. This interface provides access to all datasets except those containing sensitive or restricted information.
- An open data download interface, enabling the extraction of datasets to support research, analysis, and other scholarly activities.

Additionally, several specialized public data resources and platforms designed to promote public engagement have been developed based on HUMMA:

- garamantas.lv – initially developed as the digital archive of the Archives of Latvian Folklore, this platform was later supplemented with materials from the University of Latvia's Faculty of Humanities, the Latgale Student Center / association "Volūda", Riga Technical University Liepāja, Rēzekne Academy of Technology, and published folklore materials (Reinsone and Laime, 2022);
- Simtgades burtnieki (<http://lv100.garamantas.lv>) and Simtgades talka (<https://lfk100.garamantas.lv>) – subdomains of garamantas.lv synchronized with HUMMA, designed for public involvement in transcribing manuscripts and audio files (Reinsone, 2018; Reinsone, 2020; Pērle-Sīle et al., 2022);
- literatura.lv – Latvian Literature Digital Resource, featuring a database of authors and literary works from the 16th century to the present (Eglāja-Kristsons and Graudiņa, 2024);
- womage.lv – data on women's intellectual activism as a driving force for social and cultural change in Latvia up to World War II;
- izrades.lv – a portal for current and historical Latvian theatre performances, created in collaboration with ILFA;
- ltds.lv – the website of the Latvian Theatre Workers' Union, created in collaboration with ILFA (sections "Spēlmaņu nakts" and "Members");
- iesaieties.lv – a digital platform for public science initiatives in the humanities and arts, where opportunities for public engagement with ILFA's digital resources are gathered (Gailīte et al., 2024, Darģis et al., 2024, Pērle-Sīle et al., 2024).

The current content of HUMMA includes 28 data repositories, comprising 5,786 different data collections. These collections include 1,129,135 digitized files, 474,764

textual units with their associated metadata, 19,200 audio and video recordings with accompanying metadata, 39,936 illustrations and their metadata, 63,300 profiles of literary works, 3,464 theatre performance profiles, 306 cultural award profiles, 36,867 person profiles, 6,692 organization profiles, and 28,031 place profiles. Furthermore, a part of the text corpora developed within the HUMMA digital infrastructure has been integrated into external repositories to enhance accessibility and facilitate wider use. Examples of such integrations include the Clarin-LV repository and the National Corpus Collection (Skadiņa et al., 2022; Dargis and Saulīte, 2024).

5. Conclusion

The creation of a unified framework, while time-consuming and not always smooth, has proven to be beneficial in enabling interdisciplinary collaboration and opening up opportunities for researchers and other users to uncover unexpected connections or links within data collections of different research disciplines. Another critical aspect is ensuring sustainability, which is more likely to be achieved for a substantial resource maintained as an institutional responsibility. In this context, research projects utilizing the HUMMA infrastructure tend to provide additional contributions in terms of human resources, development, and funding, but the digital infrastructure itself does not rely solely on these projects for its maintenance or continuity. The ongoing development of HUMMA is being undertaken as part of the State Research Program “Digital Humanities” project, “Towards Development of Open and FAIR Digital Humanities Ecosystem in Latvia” (DHELI, 2022–2025). This initiative focuses on implementing new functionalities, including the creation of automated statistical annotations for folklore collections based on metadata. Additionally, the project aims to develop a text comparison tool and a specialized interface for generating text corpora and conducting text analysis, thereby enhancing HUMMA’s services for humanities research.

Acknowledgments

This paper was supported by the project “Towards Development of Open and FAIR Digital Humanities Ecosystem in Latvia” (No. VPP-IZM-DH-2022/1-0002) of the State Research Program “Digital Humanities”, funded by the Latvian Council of Science.

References

- Barats, C., Schafer, V., Fickers, A. (2020). Fading Away... The Challenge of Sustainability in Digital Studies. *Digital Humanities Quarterly*, 14(3).
- Dargis, R., Saulīte, B. (2024). Korpuss.lv – a Versatile Platform for Digital Humanities. *Baltic J. Modern Computing* 12(4), 636-645
- Dargis, R., Znotiņš, A., Auziņa, I., Saulīte, B., Reinsone, S., Dejus, R., Kļavinska, A., Grūzītis, N. (2024). BalsuTalka.lv – Boosting the Common Voice Corpus for Low-Resource Languages, in Calzolari, N., Kan, M.-Y., Hoste, V., Lenci, A., Sakti, S., and Xue, N. (eds), *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pp. 2080–2085.

- Edmond, J., Morselli, F. (2020). Sustainability of Digital Humanities Projects as a Publication and Documentation Challenge. *Journal of Documentation*, 76(5), 1019–1031. <https://doi.org/10.1108/JD-12-2019-0232>
- Eglāja-Kristsonsone E., Graudiņa, E. P. (2024). Exploring the Role of Literatura.lv in Enhancing Knowledge through Digital Humanities. *Baltic J. Modern Computing* 12(4), 526–534
- Eglāja-Kristsonsone, E. (2021). Digital Resource Literatura.lv: An Introduction, *Colloquia* 47, 149–159. <https://doi.org/10.51554/Coll.21.47.08>.
- Gailīte, E., Reinsone, S., Ļaksa-Timinska, I. (2024). Promoting Digital Citizen Science for Humanities and Cultural Heritage: iesaisties.lv. *Baltic Journal of Modern Computing* 12(4), 468–476.
- Pērle-Sīle, G., Reinsone, S. (2022). Manuscript Transcription as a Tool for Deep Reading and Teaching of Folklore, *DHNB 2022, CEUR-WS Proceedings Vol-3232*, pp. 390–400.
- Pērle-Sīle, G., Krūze, A., Pleiko Z. (2024). Learning Folk Songs through the Game. *Baltic J. Modern Computing* 12(4), 585–591.
- Reinsone, S. (2018). Participatory Practices and Tradition Archives, in Harvilahti, L., Kjus, A., O'Carroll, C., Österlund-Poetzsch, S., Skott, F., and Treija, R. (eds), *Visions and Traditions: Knowledge Production and Tradition Archives*, Suomalainen tiedeakatemia, Helsinki, pp. 279–296.
- Reinsone, S. (2020). Searching for Deeper Meanings in Cultural Heritage Crowdsourcing, in Hetland, P., Pierroux, P., and Esborg, L. (eds), *A History of Participation in Museums and Archives*, Routledge, Abingdon, pp. 186–210.
- Reinsone, S., Laime, S. (2022). The Digital Archive of the Archives of Latvian Folklore *garamantas.lv*: Prehistory and Development (in Latvian). *Letonica* 47, 52–69. <https://doi.org/10.35539/LTNC.2022.0047.S.R.S.L.52.69>
- Skadiņa, I., Auziņa, I., Dargis, R., Voitkāns, A. (2022). CLARIN Language Resources and Technology Infrastructure for the Humanities and Social Sciences (in Latvian). *Letonica* 47, 312–327. <https://doi.org/10.35539/LTNC.2022.0047.I.S.I.A.R.D.A.V.312.327>
- Terras, M. (2022). Digital Humanities and Digitised Cultural Heritage, in O'Sullivan, J. (ed.), *The Bloomsbury Handbook to the Digital Humanities*, Bloomsbury, pp. 255–266. <https://doi.org/10.5040/9781350232143.ch-24>
- Tucker, J. (2022). Facing the Challenge of Digital Sustainability as Humanities Researchers. *Journal of the British Academy*, 10, 93–120. <https://doi.org/10.5871/jba/010.093>