Sustainable, Usable and Visible Digital Cultural Heritage: Twinning for Excellence (DIGHT-Net)

Data Management Plan

Version 1.0 (March 2025)

Duration: 1 October 2024–30 September 2027 (36 months)

Lead Institution: Tallinn University

Partners: University of Bologna, University of Amsterdam, University of Turku

1. Data Summary

Will you re-use any existing data and what will you re-use it for?
The project will use existing data from the Juri Lotman Semiotics Repository and relevant digitized archival materials. Additionally, some external data sources, including Umberto Eco's materials, will be referenced but not incorporated into the project's openly available datasets due to restrictions imposed by Eco's family and the University of Bologna.

Data from the Juri Lotman Semiotics Repository will be used for:
- Digitization and long-term preservation of cultural heritage materials.
- Scholarly analysis and interpretation.
- Public access and dissemination where legally permitted.

Data from Umberto Eco's family will be used strictly for scholarly research and will not be shared openly, in compliance with legal restrictions and management policies handled by the University of Bologna.

What types and formats of data will the project generate or re-use?

- Textual data (PDF, DOCX, TIFF, XML) – archival materials, metadata descriptions, and scholarly analysis
- Audio-visual data (MP3, MP4, TIFF, JPEG, PNG) – digitized materials, recorded seminars and workshops
- Metadata (XML, CSV) – structured metadata for cataloguing and referencing
- Research reports, guidelines, and theoretical outputs (DOCX, PDF)

What is the purpose of the data generation or re-use and its relation to the objectives of the project?
The data will support the creation of a Digital Archive of Juri Lotman, contribute to developing a semiotic theory of digital cultural heritage, and enable the sustainability of digital cultural heritage research.

What is the expected size of the data that you intend to generate or re-use?
The estimated size is approximately 10 TB, considering high-resolution scans, audiovisual content, and extensive metadata records.

What is the origin/provenance of the data, either generated or re-used?

- Juri Lotman's materials (Tallinn University's Juri Lotman Semiotics Repository and University of Tartu Library; both institutions allow us to do scholarly analysis, digitise and publish the relevant materials)
- Umberto Eco's materials (restricted access, managed by Eco's family in cooperation with the University of Bologna; we are allowed to work with that data for scholarly analysis)
- Research outputs generated by the project's consortium, consisting of Tallinn University, University of Amsterdam, University of Bologna, and University of Turku

To whom might your data be useful ('data utility'), outside your project?

- Researchers in digital cultural heritage, semiotics, and archival studies
- Digital humanities scholars
- Librarians and digital archive curators
- Policy-makers in cultural heritage preservation

2. FAIR Data

Making data findable, including provisions for metadata

- All data will be assigned persistent identifiers (DOIs where applicable). Except the data belonging to Eco and some specific datasets related to Lotman that will be listed here when identified.
- Metadata will follow the Europeana Data Model (EDM) and Dublin Core standards.
- Keywords related to semiotics, cultural heritage, and digital archiving will be included.

Making data accessible

- Data will be stored in trusted repositories:
    - Digitized materials of Lotman archives will be published on ETERA – the digital library of the Academic Library of Tallinn University.
    - Open data from the project will be published on DataDOI.
    - Sensitive data from Lotman's archive will be stored by the Lotman Semiotics Repository.
    - Eco's materials will be stored and managed by the University of Bologna, which will manage the data and access to it together with Eco's family.
    - Access to sensitive data will require submitting a CV and motivation letter, which will be evaluated by the data custodians.
- Data will be openly available, except for some third-party materials and sensitive information from Lotman's archive, as well as Eco's materials, which require direct contact with Eco's family and a formal application process.
- Access to Eco's restricted materials requires submission of a CV and motivation letter, followed by assessment by the copyright holders.

Making data interoperable

- Data formats will comply with standard metadata schemas (XML, RDF, CSV).
- Interoperability with Europeana and other international digital heritage repositories will be ensured.

Increase data re-use

- Open-access licensing (CC BY-NC 4.0 for consortium-owned data, CC0 for metadata).
- Data documentation, including methodology descriptions and re-use guidelines.
- Long-term accessibility will be maintained beyond the project's duration.

3. Allocation of Resources

What will the costs be for making data or other research outputs FAIR in your project?

- Storage and maintenance: Estimated €10,000 over the project duration.
- Data curation and metadata generation: €5,000.

How will these be covered?

- Costs are included in the Horizon Europe grant as part of research data management funding.

Who will be responsible for data management in your project?

- PI Marek Tamm (Tallinn University, coordinator) and dedicated data stewards at partner institutions (University of Bologna, University of Amsterdam, University of Turku).

How will long-term preservation be ensured?

- Digitised materials from Lotman's archive will be stored in Tallinn University's digital infrastructure (ETERA) with mirrored backups. ETERA does not have a time restriction.
- Data related to scholarly work will be preserved on DataDOI at least for 10 years after the end of the project.
- High-quality archival formats (e.g., TIFF, PDF/A, XML) will be used, with periodic reviews to ensure their relevance.
- Selected materials will be archived in European digital preservation initiatives (Europeana).
- Juri Lotman Semiotics Repository at Tallinn University will oversee the maintenance and sustainability of preserved materials from Lotman's archive beyond the project's funding period. The task to oversee the maintenance and sustainability of materials from Eco's archive is secured by the University of Bologna in cooperation with Eco's family.

4. Data Security

What provisions are or will be in place for data security?

- Data will be stored securely on Tallinn University's IT infrastructure, with access restricted to authorized project members only.
- Security measures include:
  - Secure cloud storage (Google Drive) managed by Tallinn University.
  - Guidelines for secure data access, including the prohibition of data transfer to unauthorized devices or networks.
  - Encrypted data storage and regular backups.
  - Secure servers with restricted access for sensitive data.
  - Regular backups and encrypted storage for restricted materials.
- Informed consent forms will be collected for data sharing and long-term preservation, detailing how data will be processed, stored, and archived.

Will the data be safely stored in trusted repositories for long-term preservation and curation?

- Yes, the digital library of the Academic Library of Tallinn University, ETERA, will be used for the digitised materials of Lotman's archive.
- Data related to scholarly work will be preserved on DataDOI.

5. Ethics and Legal Compliance

Are there any ethics or legal issues that can have an impact on data sharing?

- Personal correspondence from third parties in the archives of Lotman and Eco will be treated with respect to privacy regulations.
- Eco's materials are subject to legal and contractual restrictions; users must apply for access. Tallinn University's legal team was involved to ensure that all restrictions will be properly respected.

Will informed consent for data sharing and long-term preservation be included in questionnaires dealing with personal data?

- Yes, all contributors and participants in the workshops and seminars will be required to sign informed consent forms for data sharing. These informed consent forms will include: what data is collected, how it will be processed, stored, and archived.

6. Other Issues

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management?

- The project follows Tallinn University's Research Data Management policies and the European Code of Conduct for Research Integrity.
- Compliance with Estonian Research Council (ETAg) guidelines on open science and data management.

Conclusion

The DIGHT-Net project is committed to open data principles while respecting legal and ethical constraints on certain materials. The DMP ensures sustainability, accessibility, and compliance with FAIR principles.