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

# Concept, Strategy, and Action plan for building the Twinned Digital Archives of Lotman and Eco

**Version 1.0 (March 2026)**

Prepared by Yan Asadchy, Merit Maran, and Marek Tamm

**Lead Institution**: Tallinn University

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

This document presents the concept, strategy, and action plan for building the twinned digital archives of Juri Lotman and Umberto Eco as a shared yet distributed infrastructure for the study of European intellectual heritage. The initiative is based on the premise that the two archives should develop as distinct yet interoperable environments capable of supporting preservation, scholarly research, and cross-archive exploration. Within this broader framework, the Lotman Digital Archive prototype constitutes the first implementation step, establishing the conceptual foundations, metadata model, and technical principles that will guide the future development of the Eco archive.

The document introduces the conceptual foundations of the Lotman Digital Archive, outlines the strategic principles guiding its design and interoperability, and presents a phased implementation plan for developing the infrastructure, digitising materials, and building an exploratory interface.

# I. CONCEPTUAL FRAMEWORK

The Lotman Digital Archive is conceived both as a practical digital infrastructure and as an intellectually informed model of archival knowledge organisation. Its development is guided by a set of core objectives and by a conceptual understanding of the archive as an active semiotic environment rather than a passive repository. The following sections outline these two interrelated dimensions.

## 1. Main objectives

The Lotman Digital Archive is designed to pursue a set of interrelated objectives, ranging from preservation, access, and research to scalability and applicability across different contexts. These objectives were created to reflect the broader conceptual ambition of the archive as a dynamic semiotic environment. More specifically, they translate the foundational principles of the archival continuum, semiotic processing, and intellectual ecology, understood as a model of the broader interconnected intellectual environment, into concrete, actionable requirements for software development guiding each phase of the project. Together, they define what the archive must accomplish in its immediate form as a prototype while remaining oriented toward the long-term goal of a twinned, interoperable infrastructure connecting the intellectual legacies of Juri Lotman and Umberto Eco. Thus, the Lotman Digital Archive should:

- preserve and digitise key archival materials
- make them accessible under FAIR principles, including, where possible, the provision of downloadable open-access datasets
- support scholarly research and teaching
- enable computational and AI-assisted exploration
- serve as a scalable model for twinned or networked intellectual archives

## 2. Conceptual model: Archive as a dynamic semiotic environment

The Lotman Digital Archive is not conceived as a static digital repository, but as:

- a dynamic research and learning environment
- a semiotic laboratory for studying meaning-making in digital archival contexts
- a model for the digital transformation of intellectual heritage

Inspired by Lotman's own theory of the semiosphere (Lotman 2005 [1984]), which conceives culture as a structured yet heterogeneous semiotic whole, that is, as a semiotic space that precedes and conditions individual acts of meaning, the archive should:

- represent texts as relational units rather than isolated documents
- highlight networks (correspondence, conceptual links, translations, and institutional contexts)
- enable interpretative layering (annotations, cross-references, and metadata enrichment)

## 3. Principles of archival dynamics

The conceptual foundation of the Lotman Digital Archive rests on three interrelated principles: the archival continuum, semiotic processing, and intellectual ecology. Together, these principles define how the archive organises, mediates, and sustains its materials over time. Each carries direct design implications, translating into concrete requirements for metadata architecture, relational data structures, provenance documentation, and modes of user access that shape how the archive will be implemented. Below are the three core principles described in detail.

### *3.1. Archival continuum*

The Lotman Digital Archive operates as an archival continuum, in which archival materials continue to develop after digitisation through processes of description, enrichment, and reinterpretation. At the same time, the integrity and provenance of the original documents remain fully traceable. The archive, therefore, balances transformation and preservation, making visible the processes through which archival knowledge is organised and reinterpreted. This principle implies several design requirements:

- **Ongoing enrichment**: the archive must support continuous expansion and enrichment of archival records rather than assuming a final or fixed state.

- **Layered representations of documents**: archival objects should exist as layered entities connecting the original document, its digital surrogate, metadata descriptions, annotations, and computational enrichments.

- **Traceable provenance and transformations**: the system must preserve information about the origin, material characteristics, and historical context of documents, while documenting metadata development and editorial interventions so that archival transformations remain transparent and traceable.

### *3.2. Semiotic processing*

Digital cultural data becomes meaningful through semiotic processing. Rather than merely storing documents, the digital archive organises, contextualises, and mediates cultural materials through a series of interpretative operations. This principle translates into several design considerations:

- **Structured modelling of cultural data**: the archive must support structured metadata schemas and classification systems that allow archival materials to be organised and described in consistent and interoperable ways.

- **Contextual reconstruction**: the system must enable the reconstruction of historical, intellectual, and institutional contexts through rich metadata, multilingual descriptions, and links to external repositories and collections.

- **Relational data structures**: archival materials should be represented through relational structures that allow connections between texts, persons, concepts, and events to be explored as networks.

- **Mediated access and interpretation**: the archive must provide tools that shape how users encounter materials, including search systems, computational analysis tools, and scholarly annotation environments, and adaptable interfaces supporting different modes of access and use.

### *3.3. Intellectual ecology*

The Lotman Digital Archive is conceived as the reconstruction of an intellectual ecology rather than a simple aggregation of texts. Instead of presenting documents in isolation, the archive models the scholarly environment within which Lotman's ideas emerged, evolved, and circulated. By foregrounding the interconnectedness of manuscripts, correspondence, lectures, drafts, marginalia, and institutional networks, the archive enables users to explore the broader intellectual landscape that shaped Lotman's work, including key figures, such as Zara Mints, a literary scholar and Lotman's wife, whose work and collaboration formed an integral part of his intellectual environment. From this perspective, several design requirements emerge:

- **Relational modelling of intellectual environments**: the archive should represent texts as part of a broader intellectual environment, connecting manuscripts, correspondence, lectures, drafts, and institutional contexts through relational metadata structures that allow scholarly networks and conceptual relations to be explored.

- **Multiple entry points into the archive**: users should be able to approach the archive through different perspectives, including persons, concepts, networks, time periods, and document types, rather than through a single author-centred structure.

- **Interoperable intellectual ecosystems**: the archive should be structurally prepared for integration with other intellectual archives, such as the Eco archive, through shared metadata standards, relational data structures, and cross-archive linking.

**4. From archival principles to strategic design**

**Implementing the archival continuum:**

- Continuous enrichment → research tools enabling ongoing metadata expansion and scholarly annotation (Section II.3.3, below).
- Layered document representation → metadata architecture and interoperability framework connecting digital surrogates, metadata, and computational enrichments (Section II.2).
- Traceable provenance and transformations → persistent identifiers, structured metadata standards, and governance mechanisms for documentation and version control (Sections II.2; II.4).

**Implementing semiotic processing:**

- Structured modelling of cultural data → metadata strategy, authority files, and interoperability standards (Section II.2).
- Contextual reconstruction → multilingual metadata and repository references situating archival materials within historical and intellectual contexts (Sections II.2; II.3).
- Relational linking of materials → network visualisations and cross-document connections within the archive interface (Section II.3.2).
- Mediated access and interpretation → search tools, AI-assisted retrieval, and scholarly annotation (Section II.3.3).

**Implementing intellectual ecology:**

- Relational modelling of intellectual environments → network-based visualisations and relational metadata connecting texts, persons, concepts, and institutions (Sections II.2; II.3.2).
- Multiple entry points into the archive → timeline navigation, correspondence and topic networks, search, and AI-assisted conversational interface (Section II.3).
- Interoperable intellectual ecosystems → metadata standards, persistent identifiers, and linked data principles enabling future integration with archives such as the Eco archive (Section II.2).

## II. STRATEGIC DESIGN

This section defines the strategic principles through which the Lotman Digital Archive will be structured and organised as a sustainable, interoperable, and user-oriented digital environment. It addresses archival priorities, metadata design, technical interoperability, and modes of user access, while ensuring that the archive can later develop into a twinned or networked system without requiring conceptual or structural redesign.

## 1. Archival scope and priorities

The initial focus of the archive is on primary scholarly materials that are progressively broadened to encompass contextual and network materials before the archival model is extended toward integration with other material, such as photographs, personal items and other artifacts, and eventually with other intellectual ecosystems (e.g., the archive of Umberto Eco).

### *1.1. Core corpus*

Priority materials include:

- Manuscripts and drafts
- Correspondence
- Lecture notes
- Published works with annotations
- Photographs and audiovisual materials
- Personal items

### *1.2. Phased expansion*

- Phase 1: Core scholarly materials
- Phase 2: Contextual and network materials
- Phase 3: Applicability to other intellectual ecosystems (like Eco's archive)

## 2. Metadata and interoperability

The Lotman Digital Archive should be technically capable of becoming one node in a future twinned system without structural redesign. To achieve this, the project follows design principles that support long-term scaling and interoperability.

### *2.1. Metadata strategy*

- Structured, multilingual metadata (Estonian, English, Russian)
- Authority files (persons, institutions, places)
- Persistent identifiers

### *2.2. Interoperability*

- Alignment with international standards (e.g., Europeana Data Model)
- IIIF compatibility for images
- Linked Open Data principles

## 3. User-oriented architecture

The initial version of the system is implemented as a static, dependency-minimal application requiring no server-side runtime beyond a standard HTTP file server. This architectural decision is deliberate, as it lowers the barrier to rapidly producing and testing the prototype, while also

enabling offline use. A future version of the architecture will support connections to richer Lotman archival repositories in accordance with different access levels.

The application is structured around four tightly integrated components: a force-directed network visualisation of the correspondence, a timeline view, a metadata preview modal window, and an AI-augmented conversational interface. The conversational interface is designed to enable users to engage in guided, chat-based exploration of Lotman's work and intellectual context on the basis of digitised archival materials and their metadata. In addition, a full-text search function is implemented to support access across the corpus.

For the purposes of this prototype, all materials and metadata are stored locally and are openly accessible without authentication. Future implementation will distinguish between open access, restricted access, and a dedicated research workspace for scholars. For the purposes of this prototype, all materials and metadata are stored locally and are openly accessible without authentication. Future implementation will distinguish between open access, restricted access, and a dedicated research workspace. This distinction will follow legal, ethical, and curatorial criteria, restricting access to sensitive personal data and copyrighted third-party materials. In such cases, access may be granted to authenticated researchers under clearly defined conditions.

### *3.1. Implementation of FAIR data principles*

In the current stage of the prototype, all archival entities (persons, correspondence collections, or thematic concepts) are represented as nodes in a shared metadata schema designed to satisfy the FAIR Data Principles (Wilkinson et al., 2016).

Findability is implemented through Archival Resource Key (ARK) persistent identifiers assigned to each node, stored in the `pid` field, and displayed prominently in the preview interface. ARK identifiers conform to the ARK Alliance specification and provide globally unique, dereferenceable references suitable for citation in academic outputs. For this prototype, identifiers follow the namespace `ark:/99999/` (the EZID test namespace), which should be replaced with institutionally minted ARKs upon production deployment.

Accessibility is addressed by ensuring all metadata is embedded directly in the application data layer and rendered without authentication requirements. Each record includes the URL or name of the holding repository, enabling scholars to locate and request access to primary materials. Creative Commons licences are explicitly declared for each item: the corpus employs CC BY 4.0, CC BY-NC 4.0, CC BY-SA 4.0, and CC BY-NC-ND 4.0, reflecting the actual licensing landscape across the contributing repositories.

Interoperability is achieved by aligning metadata fields with Dublin Core (Weibel et al., 1998) and Schema.org[1] vocabularies. The `title`, `date`, `description`, `language`, `format`, `subjects`, and `license` fields map directly to their Dublin Core equivalents. The `format` field uses MIME-type strings (e.g., `application/tei+xml`, `application/pdf`, `image/tiff`, `text/plain`) to enable an unambiguous format identification. Primary materials in the corpus are encoded in TEI-XML,

[1] Collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond. – schema.org

where available, in accordance with the Text Encoding Initiative P5 guidelines[2], widely adopted across European cultural heritage institutions.

Reusability is supported by including a `relatedResources` array of ARK PIDs on each record, encoding inter-item relationships at the data level rather than only visually at the interface level. This allows the graph structure to be reconstructed from the data alone, independent of any particular rendering environment.

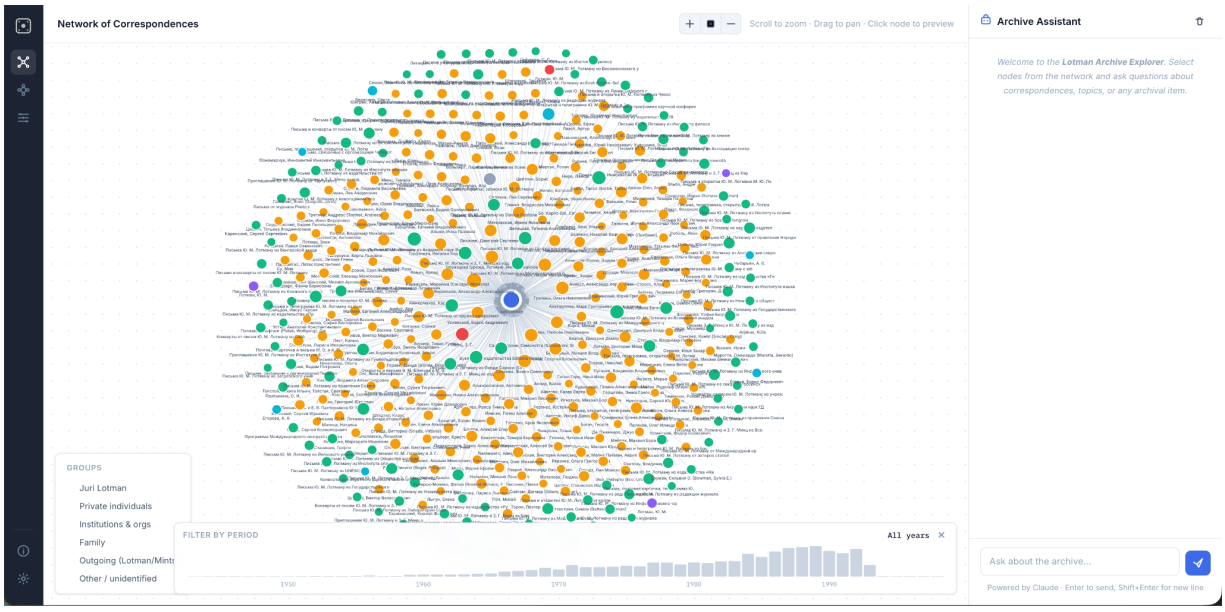
For the prototype, two exemplary datasets were created following these principles and are used for visualization and interaction until the archival data is digitised and prepared.

### *3.2. Multiple entry points*

The goal of this prototype is to provide multiple entry points for exploring the Juri Lotman archive, with plans to expand it to include other prominent figures. Each entry point provides an opportunity for users (visitors, students, scholars) to engage with the intellectual legacy of Juri Lotman at different levels of complexity through data visualisations and novel ways of interaction enabled by the AI chat interface.

The prototype interface is presented in Figure 1 below, followed by the detailed description of three main entry points.

*Figure 1. Interface of the Juri Lotman Digital Archive*




*The figure displays the Network of Correspondence view (center), one of three main entry points to interact with the digital archive. The A panel on the left allows users to switch between the three entry points, while the chat interface connects selected nodes and relevant metadata to the context of a chatbot conversation (on the right). The legend and date-range filtering (below) allow users to filter and interpret the network visualisation.*

[2] P5 Guidelines - https://tei-c.org/guidelines/p5is through, and correspondence that were digitised and/

### *3.2.1. Timeline*

One way to interact with the archival data is through the timeline view. It presents a comprehensive macroscopic view of Juri Lotman's archival legacy across more than eight decades of intellectual activity (1943–2021), aggregating over 2,500 items drawn from two principal fonds of the University of Tartu Library's digital repository. Each document with a date in its metadata is situated along the time axis, providing a bird's-eye view of Juri Lotman's scholarly activity and personal communication over the years. Each document on the timeline is clickable and provides rich metadata for inspection.

The visual grammar of the visualisation encodes two independent semantic axes. The first, shape, distinguishes authorial agency with circles representing items authored by Juri Lotman, triangles those by Zara Mints, and squares institutional or collective productions. Colour maps each item to a thematic category (e.g., published output in blue, private correspondence in amber, reception by others in purple). Users can filter the timeline by the document's origin, time period, and document type.

Crucially, the timeline does not function as an isolated view but as a chronological index that is semantically continuous with the correspondence and scholarly network graphs, allowing a researcher to triangulate a specific letter or publication through its temporal position, its relational embeddings in co-authorship and epistolary networks, and its archival origin. This integration reflects the broader methodological ambition of the prototype to support multi-perspectival analysis of a single cultural archive.

### *3.2.2. Network*

Another way users can interact with the repository is through the comprehensive network visualisation of Lotman correspondence over the years, as well as the networks of topics covered in the documents.

The example correspondence network comprises 2,506 items (F1: 1,733 correspondence and F3: 773 scholarly items) from 1940 to 2021. Weighted edges representing the documented letter exchanges between Lotman and key intellectual correspondents, drawn from archival holdings across a number of institutions (e.g., University of Tartu Library, Università di Bologna). Nodes are categorised into four scholarly groups: core (Lotman), Tartu-Moscow school, and international, reflecting the established historiography of the semiotic movement (cite your work on a). Edge weights encode the volume of digitized correspondence (number of letters), and edge labels specify the documented date range of exchange.

The network visualisation is built with D3.js v7[3], a widely adopted JavaScript library for data-driven document manipulation using a force-directed graph, allowing users to zoom in and out, pan the canvas, and move nodes, identifying clusters of correspondence based on scholarly groups. SVG rendering allows for scalable visualization across different screens and view sizes.

[3] The JavaScript library for bespoke data visualization - https://d3js.org/

The topics network comprises 13 nodes and 17 weighted edges representing Lotman's principal scholarly themes and their intellectual interconnections. Nodes are currently assigned to six thematic clusters (Published output, Process materials, Public engagement, Career & admin, Reception), with the central one focusing on semiotics as the main discipline, and other emerging concepts derived from the received texts in Lotman's bibliography [(cite repository)]. Edge labels encode the documented conceptual relationship between adjacent themes (e.g., "gives rise to", "applied to", "structural axis"), functioning as a controlled vocabulary of intellectual influence. Both datasets were modelled as dummy data for demonstration purposes; they are structured to be replaced directly with verified archival data as the project matures.

#### *3.2.3. AI-Augmented retrieval interface*

The conversational interface was implemented as a retrieval-augmented generation (RAG) pattern using Anthropic's Claude API[4] with the Claude Sonnet 4.5 model as the language-model backend. The implementation allows users to select a specific node in the network diagram and add its metadata to the AI interface's contextual window. This approach extends the classical RAG pipeline by allowing retrieval to be performed manually by the user. This design reflects two deliberate choices: it keeps the prototype dependency-free (at this stage, there is no embedding model or vector database), and it maintains scholarly agency over context construction.

When a user attaches an archival node to the chat context via the modal interface, the node's metadata record (title, PID, repository, date, description, subjects, licence, format, and related PIDs) is serialised into a structured plain-text block and prepended to the system prompt sent with each subsequent API request. The base system prompt instructs the model to act either as a specialist in Lotman's work and the Tartu–Moscow Semiotic School, or as Lotman himself, to respond concisely with citations to archival sources where appropriate, and to use Markdown for typographic structure. Up to twenty prior turns of dialogue are included in each request to maintain conversational coherence across extended sessions, subject to the model's context-window limit.

### *3.3. Research tools*

In addition to the main exploratory elements of the interface of the Lotman Digital Archive prototype, several key tools and functionalities have been identified during the design sprint workshop that took place on 26–27 February 2025 at Tallinn University:

- Full-text search (including HTR where possible).
- Named-entity recognition. Named entity recognition is performed on the materials to enhance the metadata and the experience of navigating a large collection of documents.
- Topic clustering. Emerging topics are highlighted with color and shape, making them easy to identify in related texts.

[4] Documentation - https://platform.claude.com/docs/en/home

- Cross-document linking. The modal window with the document preview includes links to the mentioned and/or related documents to facilitate navigation across scattered resources.

- Citation export. Each element of the collection contains a tool to quickly export the appropriately formatted citation for the references.

- Scholarly annotation. Each element in the collection includes a field for adding comments and annotations that support version control and reflect the archive's living nature.

### *3.4. Layered access*

The prototype uses the Tallinn University login system to determine access to specific data in the corpus. At its current state of development, it supports three types of access:

- Open-access materials for guest visitors
- Restricted access where legally necessary
- A research workspace for advanced users

### *3.5. Closing remarks*

The prototype should be understood primarily as a proof of concept for the metadata and interoperability layer (Section II.2) and as a partial demonstration of user-oriented navigation (Section II.3), with the network visualisation and conversational interface serving as two concrete interaction paradigms.

### 4. Governance and sustainability

The archive requires:

- A clear intellectual property framework
- A data Management Plan
- A long-term hosting solution
- Documentation and version control
- Training of staff and students

**Scalability principle**: The governance model should anticipate collaboration with external archives, including the Eco archive.

## III. IMPLEMENTATION PLAN

This section translates the project's conceptual and technical specifications into a working prototype and, subsequently, into a sustainable infrastructure. Prototyping is an essential practice in software development and, in this case, serves as a method for testing the archive's structure, modes of access, multiple entry points, and user experience in practice, while simultaneously

evaluating how digitised materials, metadata, and exploratory tools function together within an integrated web interface.

The prototype interface comprises three principal entry points (a timeline view, a correspondence network, and a topics network) implemented as interactive visualisations using the D3.js library and published within the Tallinn University infrastructure. An AI-augmented conversational interface, built on the Claude API, enables users to engage with selected archival metadata as contextual input, supporting guided scholarly exploration of Lotman's work. Search and browse functionalities are implemented across the corpus, and a pilot dataset is available as part of the prototype environment.

**Phase 1 – Preparation and framework**

The first phase establishes the institutional, technical, and organisational conditions necessary for implementation. It includes the definition of metadata standards and platform infrastructure, as well as preparation of the Data Management Plan, thereby ensuring that digitisation, development, and governance work in later phases can proceed on a stable basis.

***1.1. Audit and mapping***

The first phase focuses on establishing a comprehensive overview of the archival materials and the institutional conditions for their digitisation and integration. This includes mapping the fonds and series held in the Semiotics Repository at Tallinn University and in the University of Tartu Library, and identifying their scale, formats, and current level of description.

Because the archival materials are distributed across two institutions and have been organised according to partly different archival logics, special attention will be paid to differences in fonds structure, digitisation unit formation, and levels of metadata granularity. At the same time, a rights assessment will be carried out to determine which materials can be made openly accessible, which require restricted access, and which may involve additional permissions from copyright holders.

While Tallinn University holds the copyright to Juri Lotman's intellectual property, the archive also contains third-party materials, such as correspondence, manuscripts, publications, and other documents, as well as materials that may require restricted access due to sensitive content. Prioritisation criteria will therefore be defined on this basis, taking into account scholarly relevance, technical feasibility, access conditions, and the value of particular materials for future integration into the twinned archive environment.

***1.2. Technical planning***

**Platform selection**

The digital archive will rely on trusted institutional repository infrastructures already used by the partner institutions. Digitised materials from the Lotman archives will be stored and published through ETERA, the digital library of the Academic Library of Tallinn University, with parallel availability in ADA, the digital archive of the University of Tartu Library. Bibliographic records will be linked to the shared Estonian library catalogue ESTER, ensuring discoverability within the

national library infrastructure. The repository environment will support open access where legally permitted while allowing restricted access for materials subject to copyright or privacy constraints.

**Metadata schema definition**

A key task in this phase is to ensure compatibility between materials and metadata produced in Tallinn and Tartu, while also preparing the archive for future twinning with the Eco archive. Since the archival materials in both institutions are already catalogued in the shared Estonian library system ESTER, which follows widely used metadata standards and interoperability practices, the integration of the collections into a common digital environment is expected to be technically feasible. Metadata standards will follow the Europeana Data Model and Dublin Core to ensure interoperability with international digital heritage infrastructures.

**Security and data governance**

In accordance with secure software development principles (OWASP ASVS, Level 1; OWASP 2021)[5], a number of controls were implemented. First The Anthropic API key needed for the functioning of the AI Interface is stored exclusively in the browser's `sessionStorage`, which is scoped to the active tab session and automatically cleared when the tab is closed. The key is never written to `localStorage`, never embedded in source files, and never transmitted to any endpoint other than `api.anthropic.com` over HTTPS. The application prompts the user to supply the key via a dedicated settings dialog on first use; a format-validation regular expression (`/^sk-ant-[a-zA-Z0-9\-_]{20,}$/`) is applied client-side before the key is accepted.

Second, all user-supplied chat input is transmitted to the API as plain text data and is never interpolated into the Document Object Model (DOM) as markup. Error messages returned from failed API calls are passed through a regular expression that redacts any string matching the API key pattern (`sk-ant-...`) before display or console logging, preventing accidental credential exposure in error states.

**Data management plan**

The phase also includes preparation and implementation of the project's Data Management Plan (DMP), which defines procedures for data storage, documentation, access, and long-term preservation in accordance with FAIR data principles. Digitised materials from the Lotman archives will be preserved in the Tallinn University digital library ETERA, while research outputs and open datasets generated by the project will be deposited in trusted repositories such as DataCite-supported or institutional research data repositories. The DMP also specifies procedures for managing restricted materials, including third-party correspondence and sensitive archival content.

[5] OWASP Application Security Verification Standard (ASVS) – https://owasp.org/www-project-application-security-verification-standard

**Phase 2 – Digitisation**

The second phase focuses on the systematic digitisation of the archival materials held in Tallinn and Tartu. Digitisation is carried out by the Tallinn University Centre for Digitisation in cooperation with the University of Tartu Library. Because of the material diversity and the physical characteristics of the archival holdings, the digitisation process must largely be handled manually rather than through automated workflows.

### *2.1. Digitisation*

Preservation copies are produced in TIFF format, using high-resolution scanning and colour settings suitable for long-term preservation. Access and archival copies are generated in PDF/A format. Where the quality of the material allows, optical character recognition and handwritten text recognition will be applied in order to improve the searchability and usability of the digitised corpus. The digitised materials will be made available through the respective institutional environments, including ETERA and ADA, and linked to bibliographic records in ESTER.

### *2.2. Metadata creation*

Metadata creation accompanies the digitisation process and includes both technical and descriptive metadata. Technical metadata documenting the scanning process is generated automatically and includes information such as file format, resolution, colour depth, hardware and software used, and file creation data. Descriptive metadata is added in accordance with established cataloguing practices and includes fields such as title, author, date, language, format, keywords, and repository reference. Further metadata harmonisation and enrichment may be undertaken in later stages of the project.

**Phase 3 – Development of prototype**

This phase focuses on translating the project's conceptual and technical principles into a working prototype. Prototyping serves here as a method for testing the archive's structure, modes of access, and user experience in practice, while also making it possible to evaluate how digitised materials, metadata, and exploratory tools function together in an integrated environment.

### *3.1. Prototype interface*

Core tasks include:

- Design and development of a user interface comprising three principal entry points: timeline, network of correspondence, and network of topics.
- Implementation of interactive visualisations of the digitised corpus using D3.js libraries.
- Development and publication of a first public prototype within the Tallinn University infrastructure.

- Integration of a chat interface enabling users to interact with the Claude API on the basis of selected archival metadata as contextual input, thus supporting guided exploration of Lotman's work through digitised materials.
- Implementation of search and browse functionalities across the corpus.
- Publication of a pilot dataset online as part of the prototype environment.
- Further refinement and consolidation of metadata guidelines in light of the practical requirements of interface design, corpus visualisation, and user interaction.

**Phase 4 – Enrichment and research integration**

This phase is dedicated to enriching the prototype and integrating it more fully into scholarly research practices. It combines computational methods, expert-driven annotation, and user testing in order to enhance the archive's analytical potential, improve its usability, and support its development as a research infrastructure for the study of Lotman's work and related intellectual contexts.

***4.1. Computational enrichment***

This phase contains tasks related to the integration of computational text analysis methods to assist scholarly professionals in their work with the archive:

- Entity extraction using the Structured Entity Extraction (SEE) framework (Morejón et al., 2025);
- Topic modelling using BERTopic modelling (Grootendorst, 2022);
- Network visualisation using D3 JS framework, the ability to capture and save visualisations in image accessible format.

***4.2. Scholarly annotation layer***

This phase contains tasks related to the integration of the annotation layer that will help users to discover, update and properly cite the archival material:

- Expert annotations in the form of comments and edit suggestions;
- Thematic pathways (e.g., "Semiosphere", "Cultural Memory");
- Linked bibliographies.

***4.3. User testing***

The goal of this prototype is to provide new ways of interaction with the digital archive. To understand which interactions are effective and to promote new ways of information retrieval, hypothesis generation, search, and answering research questions, it is critically important to test the proposed solution with future users.

As described in greater detail above, the prototype serves three main user groups:

- Researchers
- Graduate students
- Heritage professionals

Testing with each user group will be carried out at Tallinn University according to a clearly defined protocol. Each group will take part in a shadowing exercise facilitated by a member of the project team, who will observe the sessions, take notes, and provide assistance where necessary without influencing participants' decisions. All participants will receive onboarding materials, including information about the test, a consent form, and the test scenarios they will be asked to complete.

Observations will be systematically documented during the study, and participants will complete an offboarding questionnaire designed to evaluate both the qualities of the interface and their overall experience of using it. No incentives are planned for participants. The results of the testing phase will be synthesised in an evaluation report and used to guide subsequent design and development decisions, including the refinement of interface mock-ups and user study materials for later iterations.

**Phase 5 – Consolidation and scaling**

The final phase focuses on transforming the prototype into a stable, documented, and scalable digital archive infrastructure. At this stage, the aim is no longer only to demonstrate functionality, but to consolidate workflows, stabilise the technical environment, and prepare the Lotman archive for future integration into a twinned archival system with the Umberto Eco archive. This phase therefore, links sustainability, interoperability, and dissemination.

***5.1. Consolidation of infrastructure and workflows***

In this phase, the technical and organisational components developed during the earlier stages are brought together into a stable operational framework. The prototype interface, metadata workflows, digitisation outputs, and enrichment pipelines are reviewed and standardised in order to ensure that the archive can continue to function beyond the project period. This includes documenting all major workflows, clarifying institutional responsibilities, and establishing procedures for future maintenance and updates.

Particular attention will be paid to the relationship between the archival repositories (ETERA, ADA, ESTER) and the exploratory research interface. The goal is to ensure that the user-facing environment remains clearly connected to the trusted institutional repositories while allowing for future expansion of functions such as annotation, search, and AI-assisted retrieval.

This phase also includes the preparation of internal guidelines for long-term management of the archive, including procedures for metadata updates, ingestion of newly digitised materials, quality control, and handling of restricted-access items.

Core tasks:

- final stabilisation of the prototype and repository connections
- documentation of digitisation, metadata, and enrichment workflows
- preparation of archival and editorial guidelines
- definition of long-term maintenance responsibilities across partner institutions
- implementation of version control and change documentation procedures.

### *5.2. Sustainability and governance*

A central objective of this phase is to ensure that the archive is institutionally sustainable. The governance model developed earlier in the project will be formalised through clearly defined roles and responsibilities among Tallinn University, the University of Tartu Library, and future external partners. This includes clarifying decision-making structures for metadata changes, interface development, rights management, and future integrations with partner archives.

The sustainability framework should address technical hosting, repository preservation, software maintenance, staff training, and financial continuity. Since the archive is conceived as a long-term scholarly infrastructure rather than a temporary project output, the sustainability plan should also identify which components belong to institutional core services and which may require future project-based development.

Training materials will be prepared for archive staff, researchers, and students in order to support the continued use and development of the archive. In this way, sustainability is understood not only in technical terms, but also as the cultivation of an expert community capable of maintaining and expanding the archive over time.

Core tasks:

- formalisation of governance model and partner roles
- sustainability planning for hosting, maintenance, and preservation
- preparation of training materials for staff and users
- definition of procedures for future corpus growth and metadata enrichment
- clarification of intellectual property and rights management workflows for long-term use.

### *5.3. Twinning readiness and cross-archive interoperability*

The most important strategic objective of this phase is to prepare the Lotman archive for twinning with the Umberto Eco archive without requiring structural redesign. Building on the metadata and interoperability principles established earlier, this phase develops the practical and conceptual framework for cross-archive integration.

This work includes harmonising metadata elements that are likely to be shared across both archives, such as authority files for persons, institutions, places, works, concepts, and events;

mapping descriptive fields to a common interoperability model; and identifying the minimum shared data requirements for cross-archive discovery and linking. Persistent identifiers, multilingual metadata, and linked-data principles will be further operationalised in order to support future relational navigation between the Lotman and Eco archival environments.

A pilot model for cross-archive linking will be developed in order to test how twinned archival logic may function in practice. This may include linking shared correspondents, intellectual concepts, publication histories, translations, or parallel thematic pathways. The purpose of the pilot is to demonstrate how two distinct intellectual archives can be made interoperable while preserving their local archival specificities.

Core tasks:

- development of a metadata harmonisation framework for twinned archives
- mapping of common entities, concepts, and relational structures across archives
- preparation of interoperability documentation and shared data model guidelines
- creation of a pilot cross-archive linking model
- definition of technical and conceptual requirements for future Lotman–Eco integration.

### *5.4. Evaluation, dissemination, and community building*

The final phase also includes the consolidation of the project's scholarly and public impact. The archive and its methodology will be disseminated through academic publications, presentations, workshops, and targeted engagement with archival and digital humanities communities. Dissemination should highlight both the conceptual contribution of the archive as a dynamic semiotic environment and the practical lessons learned from building an interoperable intellectual archive.

In addition to academic dissemination, a short policy-oriented document should be produced summarising the main recommendations for digitising and connecting scholarly archives in ways that support research, interoperability, and long-term sustainability. This would strengthen the archive's value as a transferable model for other heritage and research institutions.

Core tasks:

- preparation of academic publications and conference presentations
- organisation of expert workshops and stakeholder meetings
- production of a policy brief on digitising and networking scholarly archives
- communication of project results to digital heritage and humanities communities
- identification of future funding and collaboration opportunities.

**Final overview**

The Digital Archive of Juri Lotman is conceived as a multi-layered, scalable, and future-proof infrastructure that integrates archival richness with the practicality of its exploration. It functions as a dynamic semiotic environment rather than a static repository, treating archival materials and their metadata as relational units embedded within broader networks of intellectual exchange, institutional context, time, and cultural meaning.

Its modular and interoperable architecture, grounded in FAIR data principles, Dublin Core metadata standards, persistent identifiers, and linked open data practices, ensures that the archive is prepared for expansion without requiring redesign effort. Practically, its development follows a clear, phased trajectory, moving from audit and digitisation through prototype development, computational enrichment, and user testing, and then to consolidation and long-term operation.

Most significantly, the archive is designed from the outset as one node in a future twinned system. Thus, the Digital Archive of Juri Lotman should be:

- **conceptually**: a dynamic semiotic archive
- **strategically**: modular, interoperable, research-driven
- **practically**: built in clear, phased stages
- **future-proof**: scalable into a twinned archive with Umberto Eco

This approach ensures that the Lotman archive is complete and coherent in itself, yet structurally prepared to evolve into a broader European intellectual archive infrastructure.

**References**

Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. arXiv preprint arXiv:2203.05794.

Lotman, Juri 2005 [1984]. On the semiosphere. *Sign Systems Studies* 33(1): 205–229.

Morejón, A., Navarro-Colorado, B., García-Barceló, C., Berenguer, A., Tomás, D., & Mazón, J. N. (2025). Automatic Metadata Extraction Leveraging Large Language Models in Digital Humanities. *Electronics*, 14(24), 4962.

Weibel, S., Kunze, J., Lagoze, C., & Wolf, M. (1998). Dublin core metadata for resource discovery (No. rfc2413).

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., & Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, 3(1), 1–9.