



1. Mobility Objective

- Research in the archive collection: examination of holdings, inventories, and collections relevant to Russian folk tales.
- On-site collection of information obtainable only in situ: corpus selection, cataloguing clarifications with curators.
- Research in collaboration with foreign researchers: exploratory interviews and the definition of the product features and necessary data.

2. Principal Results

- Built a curated corpus of 117 Russian tales of magic from 33 archival volumes (“ERA, Vene” collection – 16 texts; “RKM, Vene” collection – 2 texts; “TRÜ VKK” collection – 15 texts, corpus overview attached). Documented temporal and regional coverage (figures attached).
- Verified 784 pre-existing scans and coordinated +165 newly added pages via the archive.
- Completed transcription of 54 pages from closed collection and delivered draft of transcription guidelines (attached).
- Delivered an explicit sampling specification with inclusion/exclusion rules and a corpus overview.
- Produced the first quantitative overview (tables/figures) of corpus distributions by decade and by region/series, establishing a baseline for future evaluations (e.g., type trends, geographic spread).
- For tales originating from restricted collections, applied pseudonymisation to narrators and collectors using an internal mapping (kept outside the public repository); in the corpus version included here, these actors are represented only by stable pseudonymous IDs while preserving their roles and relational links.
- Provided a metadata schema and a CSV template with field definitions/examples; prepared an index with stable IDs, rights notes, etc.
- FAIR/LOD readiness (v1). Supplied a first-draft ontology mapping (DC / schema.org / SKOS; extensible toward CIDOC-CRM), plus a knowledge-graph sample (TTL) and JSON-LD examples with stable IRIs and embedded rights/provenance fields.
- Conducted 3 expert interviews; delivered a Product Brief (MVP) that consolidates user value, feature set, and research questions the tool will address.
- Produced a Permissions & Licensing plan.

3. Attachments

- [GitHub repository](#) containing all project documentation.
- Product Brief (consolidated problem statement, user value, features).
- Sampling_log (criteria and corpus-overview charts).
- Sampling_pipeline — workflow diagram.
- Corpus_A_index (117 records; call numbers, references and rights notes).
- Metadata_schema and metadata_examples.
- Ontology_overview (ontology mapping outline).



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DIPARTIMENTO
DI FILOLOGIA CLASSICA
E ITALIANISTICA

- Transcription_guidelines.
- 54 pages in phonetical transcription (.xml format)
- Permissions & Licensing policy draft.

4. Activity Summary

Week	Declared plan	On-site execution	Evidence
1	Data & Metadata Selection	EFA catalogues reviewed; sampling criteria defined, field vocabulary revised	sampling_log, corpus_a_index, metadata_schema
2	Stakeholder Interviews & Schema Refinement	3 interviews conducted; target user group and jobs-to-be-done defined	Product Brief
3-4	Data Extraction & Verification	Reprioritised to urgent manual transcriptions and coordination of scanning for the non-digitised portion of the collection; OCR pilot deferred to post-mobility	Transcription files, corpus_a_index
5	Corpus Creation & Sampling	Checked 33 volumes, found 130 tales of magic, manually check references to ATU types for tales of magic (final index)	corpus_a_index, graphs in the figures folder
6	Schema Implementation & Dashboard Prototyping	LOD schema (draft) and rights policy prepared, UX workflow documented	ontology_overview (draft), kg_sample, json-ld-example, Product Brief
7	Usability Testing & Refinement	Partial: qualitative think-aloud sessions final index check	Product Brief, corpus_a_index
8	Final documentation	Documentation and registers package ready	Readme, GitHub repository