

## **Recommendations for promoting the semantic interoperability of Reference Data between Latvian cultural heritage institutions**

DOI: <https://doi.org/10.52197/IBIV6839>

### Summary

Bibliography institute of the National Library of Latvia, 2021.

Authors: Mārīte Apenīte, Uldis Bojārs, Anita Goldberga, Maira Kreislere, Anita Rašmane, Jogita Sauka.

The recommendations have been prepared by the National Library of Latvia (NLL) as part of the [Latvian memory institution data in the digital space: connecting cultural heritage](#) project (No lzp-2019/1-0365), which is being implemented within the Fundamental and Applied Research Programme of the Latvian Council of Science.

The recommendations apply only to the controlled values required for the reproduction of cultural heritage provided by authoritative data and other value dictionaries, and stand apart from any metadata describing cultural heritage objects outside the scope of these recommendations.

When evaluating data from systems in individual sectors, their data models and standards, it must be concluded that different data formats are being used, with elements having different degrees of structurization; partly different key entities, which limit their interoperability in the utilisation of common values for controlled data elements, and other inconsistencies can be observed. The uniform and controlled utilisation of named entities for cooperation in a single platform for managing and distributing digital cultural heritage objects, specifically for the development of a Reference Data Solution, requires agreement on a new data model, data standards and formats consistent with the principles of the Semantic Web.

In the latest conceptual models of archive, library and museum data, the semantics of data elements have been assessed from the perspective of user requirements. These recommendations analyse the entities of three related-sector conceptual models (the Records in Context Conceptual Model, IFLA Library Reference Model and CIDOC Conceptual Reference Model), which are fundamental to cooperation within Latvia's

unified platform for managing and distributing digital cultural heritage objects for the uniform and controlled identification and utilisation of different names/designations in memory institution resources. The analysis is being carried out to identify common and different key entities, their main properties and characteristic relationships that could assist in the practical mapping of data elements.

Taking into account the analysis of the current situation in Latvian memory institutions and the results of comparing the key entities of the conceptual data models, the Register of Cultural Heritage Entities (KMER, which is the foundation of the Reference Data solution) should be based on authoritative data from the NLL and the National Archives of Latvia. There should also be a gradual move, within the data of other memory institutions, to the controlled utilisation of the main entities named: persons, institutions, gender, place, event, work, subject, form/genre. In terms of the new data model, it is necessary to agree on:

- which data element values to include in the KMER data model, which to take from value dictionaries;
- common rules to eliminate differences in recording (this applies to Agents, Datings and other entities and properties);
- the granularity of the data (for example, name, surname; institution, their substructures, etc.);
- data quality/reliability levels that would enable data builders to distinguish between reliable and verified data values and less reliable ones;
- the preservation of data provenance information, including identifiers, because provenance information serves to assess quality and conformity;
- the principles for mapping cross-sectoral data (concerning KMER) in the transition to entity management;
- the options for extending the data model;
- the options for developing and putting into practice linked data formats – as none of the existing systems in cultural heritage institutions directly provide authoritative data retrieval in linked data formats – so that they can be opened to more extensive utilisation, both mutually and outside the cultural sector.

Semantic technologies offer new principles for structuring information to create not only machine-readable but also machine-interpretable data – RDF, SPARQL, SKOS and other standards ensure data is structured for repeated utilisation. Linked data can be used to add contextual information and create a network of connections that better reflects knowledge in the real world. Context and connections can help researchers gain a more complete understanding of materials by encouraging groups in society to become more involved and utilise those materials. Linked data are also powerful because they are not just connected to a single system, but integrate cross-collection content, thus guiding the user not only among collections but, more importantly, between different repositories.

The recommendations also summarise examples of good practice from solutions outside Latvia (SNAC, Finto/FINNA, Wikidata, EDM, Intermi), from which the summarised recommendations follow for further co-operation between cultural heritage policy makers and implementers in Latvia. Significant changes that differ from the

usual practices of archives, libraries, museums and that need to be implemented in order to create a unified cultural heritage reference data solution with a register of entities are the entification and the removal of authority data (NAL, NLL) from closed systems and their opening to public access, in the form of value dictionaries/ontologies. In other words, the quality control of cultural heritage data should move from an authority record data model to an entity data model, where each instance of an entity and its properties (e.g., names in different languages and orthographies) and relationships or links are defined using machine-readable identifiers. Creators of descriptive metadata need only add a link to the relevant entity in their institutional core systems, and the specific information no longer needs to be rewritten. The information of the main entity classes should be kept separate from the metadata describing the cultural heritage objects, i.e. must be stored in the KMER.

Such uniform quality control means creating, from scratch, appropriate infrastructure for the management and development of cultural heritage entities, in which the type of entity can be defined, describing the properties of each entity separately, indicating its links with other entities. It must be possible to change and supplement the entity's information without losing the link with the cultural heritage objects themselves, which can be described in the core systems of memory institutions.

The recommendations are a first attempt to describe the requirements for a new cross-sectoral collaborative data model to move from records in closed systems to an open solution for key cultural heritage entities and their properties-relationships using open linked data standards. Only such a solution would open up a valuable knowledge base for use in both cultural heritage sectors and by other interested parties, as well as improving access for researchers, the education sector, tourism, the creative industries, etc. The Reference Data solution must become the basis for a platform for the digital transformation of the cultural heritage sector and the management and distribution of digital objects.

As a national aggregator, the NLL ensures the delivery of cultural heritage sector data to *Europeana*. Consequently, the transition to an entity system compatible with the *Europeana Data model* (EDM) is essential to ensure the delivery of data to *Europeana* with the least possible loss of data quality.

Translation: Juris Beņķis.

[Full text](#)