

Enabling Research of Cultural Heritage and Recent History using COURAGE Linked Data Registry

András Micsik^[0000-0001-9859-9186] and Tamás Felker^[0000-0002-8478-3756]

MTA SZTAKI, Institute for Computer Science and Control, Hungarian Academy of Sciences
micsik@sztaki.mta.hu

Abstract. The COURAGE project aims at preserving the history and heritage of dissent in former socialist countries. The items of so-called cultural opposition are curated in various collections, and COURAGE builds a registry of these collections. The registry is a Linked Data platform, and data input, editing and publishing are all done using RDF and SPARQL. We present our effort of building the registry and introduce our experiences with this unique born-RDF historical dataset.

Keywords: Cultural Heritage, History, Linked Data, Semantic Web.

1 Introduction

COURAGE (“Cultural Opposition – Understanding the CultuRal HeritAGE of Dissent in the Former Socialist Countries”) is a three-year international research project funded by Horizon 2020, the EU Framework Programme for Research and Innovation. The project is building a comprehensive online database (digital registry) of existing but scattered collections on the histories and forms of cultural opposition in the former socialist countries and thereby make them more accessible.

COURAGE aims at enabling the analysis of these collections in their broader social, political and cultural contexts. The general aim of this analysis is to allow for the expanded outreach and increased impact of the collections by assessing the historical origins and legacies of various forms of cultural opposition.

The collections discovered and described by the project also form the part of European cultural heritage. However, the complete digital preservation of materials from more than 1000 collections is unfeasible, so the project has chosen to focus on the collections owning and curating this cultural material.

In the registry we need to describe the history of the collections, the key contributors and stakeholders, and some of the most significant items in the collection. Furthermore, the textual descriptions need to be multilingual, usually in English and the local languages, but altogether we need to handle translations in 15 languages, as the research covers the area of all former European socialist countries, and some countries of exile and emigration.

Based on the requirements mentioned so far, the Vitro generic Linked Data platform [1] was adapted to the needs of the registry, and a comprehensive Linked Data database

was developed. In the following we present the data schema, the dataset and the IT environment built around it.

2 The registry

The main types of the registry schema are shown in Fig. 1. Basically, there are historical entities connected to each other either directly or via temporal properties. Temporal properties have a start year and an end year (even year precision is often hard to fulfill).

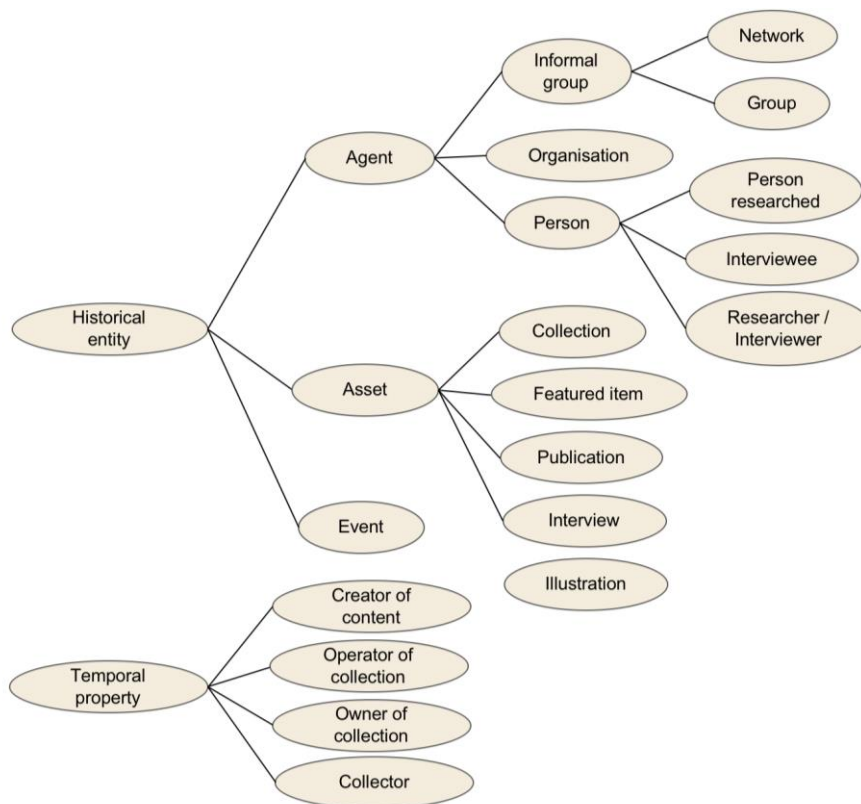


Fig. 1. Main types of the registry.

Temporal properties are needed to describe the chronology of collections. Collections are connected to events, agents and other assets, whereas featured items are significant or characteristic pieces of the collection as a sample. Most collections are described on the basis of one or more interviews, where the participants, date, location are stored, and the audio recording can also be attached, but it is usually only available for researchers. Each historical entity may have geographical locations, webpage links and illustrations associated. Illustrations are also linked data entities with several metadata

such as credits and license. There are some anonymized and private data in the registry, and the licenses often do not permit the re-use of illustrations outside the registry.

Fig. 2 gives an overview of the main connections among our RDF classes. Collections have significant events and featured items, while all of these have associated agents such as creators, key actors, founders, etc. An agent can be a group or organization or a single person. Groups and organizations may also be connected to some of their key members but listing all such members is not feasible.

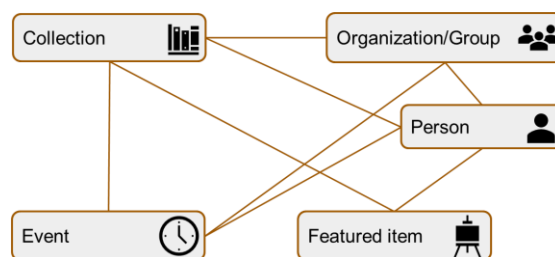


Fig. 2. Connections among main types in the registry.

Researchers use the Vitro user interface to add or delete object and data properties in the dataset. The COURAGE ontology directly defines the editors' interface where object and data properties are clearly represented and can only be changed or deleted individually. When adding text about an item, researchers enter the text and select the language. Then, they can enter the translation of the text in another language as a new data property. When connecting two entities with an object property, they can either select from existing entities in the property range class or create a new entity with a new label. Labels should also be given in several languages: in English and in the language(s) of the country the entity belongs to.

We had to add administrative metadata to support the necessary quality control in this project where more than 100 researchers are editing the knowledge base. Only after a thorough national and global historic and linguistic review may the new items appear on our published pages. Since we have no clear boundaries of quality control targets, the main classes shown in Fig. 2 serve as publication points. This process is supported by many custom administrative properties in the dataset, such as quality control status, last editor, last editing time, last review time, respective researcher task, etc.

The schema development was quite evolutionary, as historians or social scientists did not know exactly in advance what kind of data would be useful to collect. We had to adapt and extend the schema almost each month of the project. Due to the project nature, there are mostly domain specific properties in the schema, not found yet in any other related ontology. The COURAGE ontology is maintained in OWL and available for download [2], but it yet needs to be extended with class and property mappings to other vocabularies. We would use subproperty inferencing and rules in our platform, but Vitro did not support these at project start. Our aim is to provide mappings to several widely used metadata formats and RDF schemas and offer a wide range of export functionality. We also plan to implement mappings to CIDOC-CRM [3] as it has some interesting uses for describing our recent history [4].

3 Re-using the registry content

Fig. 3 summarizes the IT infrastructure of the project with the main project outcomes on the top including learning material for secondary schools and universities, virtual exhibitions, portals in 15 languages and various scientific publications. All these are using the registry as a backend either via actionable links or via SPARQL queries to the registry. The registry has a lightweight connection to a Redmine server [5] linking registry items with Redmine issues to enable task management.

The registry has a public view which is freely browsable and searchable [6]. Each published entity has a schema.org description. We are registering DOIs for the quality-controlled descriptions of entities at DataCite using DataCite Metadata Schema. We also plan to implement Europeana EDM mapping. Furthermore, we implemented HTML and Word dumps of ‘graph neighborhood’ of main items suitable as raw data for publications or blog entries. License restrictions, anonymity requests and privacy make it difficult for us to provide dataset dumps and an open SPARQL endpoint, but we are planning to do both soon and allow for more re-use of the COURAGE RDF dataset.

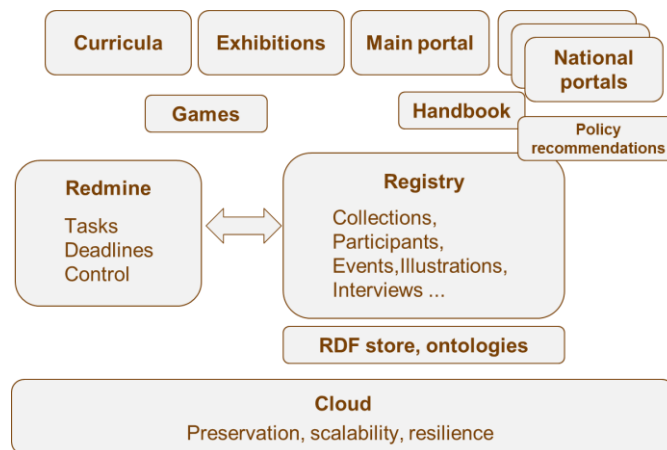


Fig. 3. The IT environment and outcomes of the COURAGE project.

References

1. VIVO/Vitro Homepage. <https://wiki.duraspace.org/display/VIVO>
2. COURAGE ontology. <http://cultural-opposition.eu/rdf/courage.owl>
3. Crofts, N., Doerr, M., Tony, G.: The CIDOC Conceptual Reference Model: A standard for communicating cultural contents. *Cultivate Interactive* 9. (2003)
4. Leskinen, P., Koho, M., Heino, E., Tamper, M., Ikkala, E., Tuominen, J., Mäkelä, E., Hyvönen, E.: Modeling and Using an Actor Ontology of Second World War Military Units and Personnel. In: d'Amato C. et al. *The Semantic Web – ISWC 2017*. LNCS, vol 10588. pp. 280-296. Springer International Publishing AG (2017)
5. Redmine project management web application. <https://www.redmine.org/>
6. COURAGE Public registry. <http://cultural-opposition.eu/registry/>