

## **The EDPOP VRE: a virtual research environment for collaborative querying, collecting and annotating works of European popular print**

Jeroen Salman, Julian Gonggrijp, Tijmen C. Baarda (Universiteit Utrecht)

We demonstrate a system to collaboratively query, collect and annotate records from libraries and cultural heritage institutions from around Europe. This system is set up as a virtual research environment (VRE) and has been created as part of an international research project that asks the question to what extent popular print culture in the period 1450–1900 was a European phenomenon with a transnational infrastructure [1]. The main goal of this project is to study popular printed matter in all its diversity, such as almanacs, newspapers, pamphlets, prose novels, songs and penny prints, from both a national and a European perspective. Many stories and songs travelled throughout Europe for long periods of time through translations and adaptations. Adopting a European perspective provides a better understanding of the origin, context, and popularisation strategies of this material. It is important to consider not only the texts and images, but also the cross-border networks of authors, translators, publishers, booksellers and itinerant traders who were involved in the production and distribution of popular printed matter.

To study large numbers of materials from different libraries, unified catalogues such as the Universal Short Title Catalogue (USTC) and the Heritage of the Printed Book Database (HPB) are very helpful, but researchers still have to resort to a multitude of different library catalogues. In our VRE, users can search and work with multiple catalogues from libraries and cultural heritage institutions, as well as a number of databases created by researchers. Users can add the records they find to private or public collections, and add annotations to the records as a whole or to their individual fields, and classify the records according to a glossary of genres [2]. While the VRE was created with European popular print culture in mind, the system is flexible and can be extended to support other library and cultural heritage catalogues, as well as entire different record types, such as musical compositions or movies. This flexibility, which makes the VRE useful for other projects as well with minimal changes, is largely due to the application of linked open data.

In our approach, we did not aim to create a unified catalogue from different sources, but we embrace the fact that biographical and bibliographical data is available in a multitude of different formats and sources. When a user of the VRE enters a query for a certain library catalogue it is performed on the fly, after which the records are presented in a normalized way. Thanks to this normalization, users can put records of different catalogues together in the same collection. The original records of the source catalogue are always quickly available for consultation.

Once a user has created a collection and added records to it, they can flexibly search and filter the records within the collection based on any field. Annotations on the records are also taken into account. The same record can be added to multiple collections at the same time. The VRE is also able to deduplicate a record if it appears in multiple catalogues, and to update records when the data from the source catalog have changed.

Architecturally, the VRE consists of a web application created using the Django and Backbone frameworks, which communicates with a Python library that is responsible for fetching and normalizing data from library sources on the fly. Data is internally stored as linked data in RDF format. The data model relies on the Web Annotation Vocabulary, which is a W3C standard, as well as two RDF ontologies that are published independently of the

VRE. The first ontology provides a generic way to describe curated collections. The second ontology is used to describe bibliographical and biographical records in the normalized way that is used in this project. The Python library can also be used on its own as a means to query various library catalogues in a unified way for use in other projects. Both the VRE [3], the accompanying Python package [4] and the two ontologies [5][6] are available under permissive open source licenses.

In our demonstration, we first show how the VRE helps with collecting and collaboratively studying European popular print that is spread over a large number of different catalogues. After that, we dive deeper into the technical aspects of the software, including a discussion of the reasons to use linked data for internal storage, and our efforts to keep the original catalogue data available at all times even after normalization to a unified format.

[1] <https://edpop.wp.hum.uu.nl/>

[2] <https://edpop.wp.hum.uu.nl/glossary/>

[3] <https://github.com/CentreForDigitalHumanities/EDPOP>

[4] <https://github.com/CentreForDigitalHumanities/edpop-explorer>

[5] <https://github.com/CentreForDigitalHumanities/edpop-collection-ontology>

[6] <https://github.com/CentreForDigitalHumanities/edpop-record-ontology>