

EXPANDING THE SCOPE. CREATING AND VALORIZING A LOCAL EXTENSION OF ROR AS INSTRUMENT TO IDENTIFY ADDITIONAL ACTORS IN FLEMISH (SSH) RESEARCH

Peter Aspeslagh

CENTRE FOR R&D MONITORING (ECOOM), UNIVERSITY OF ANTWERP, BELGIUM

Keywords: *organization database, author affiliation data, identifiers, SSH, interoperability*

International organization databases, like the Research Organization Registry (ROR), are essential tools to identify unique organizations for a wide array of purposes. Experience with author affiliation data enrichment pointed to the opportunities that a local expansion of an international database can provide.

In 2019, a parameter measuring international collaboration was added to the Flemish PRFS. Therefore, author affiliation data had to be provided for the full set of approved publications of the Flemish Academic Bibliographic Database for the Social Sciences and Humanities (VABB-SHW). As more than half of these publications are not included in the Web of Science, a large author affiliation data collection operation was launched. However, it turned out that only two thirds of the affiliated organizations could be captured by ROR. In order to ensure full coverage, new organizations had to be added to a local copy.

Extending the ROR dataset opened new avenues. It created a tool to map the full spectrum of organizations that are involved in research in Flanders, independent of their inclusion in common international databases. This instrument, the Flemish Organization Registry (FOR), is now being developed and can serve multiple purposes for different entities in the broader Flemish research community. Therefore, we want to zoom in on the content and composition of the database, the maintenance and sustainability, the interoperability of data and a governance structure.

Because the author affiliation data collection project already provided a (ROR compatible) platform for hosting organization data, the focus has quickly shifted from a technical to a *content-related perspective*. More specifically, the main sets of organizations that need to be added to the database are being defined and integrated. This process consists of several steps. First, an inventory and compilation of commonly used but not entirely interoperable Belgian organization lists was made. These lists contain, among others, a set of educational organizations (universities and colleges), major governmental entities (federal, regional and local), organizations participating

in EU funded projects etc. During a second phase, these different datasets (as well as the extended ROR database) were consolidated by both the available metadata and manual intervention. In a third step, which is currently being carried out, the extended set of organizations is integrated in the main database. It will allow the continued addition of metadata, enabling, among other features, customized categorization.

During the development of the Flemish Organization Registry, several factors were taken into account to guarantee its *sustainability and maintenance over time*. In this process, the presence of external identifiers plays a major role, as they enable quick synchronization with the most recent metadata from external sources. First, if a ROR identifier is available, automatic import of added or updated ROR data will take place. Second, if the organization disposes of a record in the Belgian Commercial Register (KBO), as most do, a synchronization is possible as well. The KBO contains precise address data, information about the nature of the organization, sector codes (NACE) etc. Additionally, EU Participant Identification Codes (PIC) and/or Wikidata identifiers can assure further enrichment.

This set of identifiers can not only be used to import the most recent metadata, but also to assure the comprehensiveness of the database. For example, if PIC's mentioned in an updated list of Belgian organizations participating in EU projects are missing in FOR, then new organizations can be identified and added to FOR in order to keep full coverage.

FOR intends to provide the data to entities that are interested to use it for enrichment of their own databases. Therefore, attention to *interoperability* is required. It is currently ensured by the already mentioned external identifiers, but these can be extended with further types of identifiers if needed for specific processes. At the moment, a first partner is harvesting selections of the already available data via a custom made API. When the a first version of the entire database will be ready, data can be released in a more extensive way. The interoperability will also allow FOR to enrich international databases with additional Flemish metadata.

In the meantime, a *governance framework* is being established. A steering committee consisting of specialists in the Flemish research community is set up to ensure the structural development of this database and its extended use by different partners.

By creating this Flemish Organization Registry, the investments in the implementation of a modification to the Flemish PRFS generate additional value. The development of a local extension to an international database like ROR shows that the general template the Research Organization Registry provides can inspire national entities to map less internationally oriented organizations. The Flemish approach is only one way to obtain a comprehensive view on organizations in a national setting; we look forward to fruitful discussions and initiatives with colleagues working on similar matter in other national contexts.