Poster

STI 2022 Conference Proceedings

Proceedings of the 26th International Conference on Science and Technology Indicators

All papers published in this conference proceedings have been peer reviewed through a peer review process administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a conference proceedings.

Proceeding Editors

Nicolas Robinson-Garcia
Daniel Torres-Salinas
Wenceslao Arroyo-Machado

OPEN ACCESS

Citation: Aspeslagh, P., Engels, T., & Guns, R. (2022). Building on ROR. Enriching and customizing multi-purpose organization databases. In N. Robinson-Garcia, D. Torres-Salinas, & W. Arroyo-Machado (Eds.), *26th International Conference on Science and Technology Indicators*, STI 2022 (sti2273). https://doi.org/10.5281/zenodo.6974621

Collection: https://zenodo.org/communities/sti2022grx/

Building on ROR. Enriching and customizing multi-purpose organization databases

Peter Aspeslagh[*], Tim Engels[*] and Raf Guns[*]

[*]*peter.aspeslagh@uantwerpen.be; tim.engels@uantwerpen.be; raf.guns@uantwerpen.be*
Centre for R&D Monitoring (ECOOM), Faculty of Social Sciences, University of Antwerp, Middelheimlaan 1, 2020 Antwerp (Belgium)

An author affiliation data collection project, processed in the context of a Flemish research evaluation parameter, led to the creation of an extended organization database. The affiliation data were coded by using international organization identifiers (GRID, later replaced by ROR). However, the project required a substantial addition of organizations as the source database did not cover the complete set of organizations mentioned on publications. This extension entails opportunities to morph into a larger organization database applicable in multiple contexts.

Using organization databases for author affiliation data collection

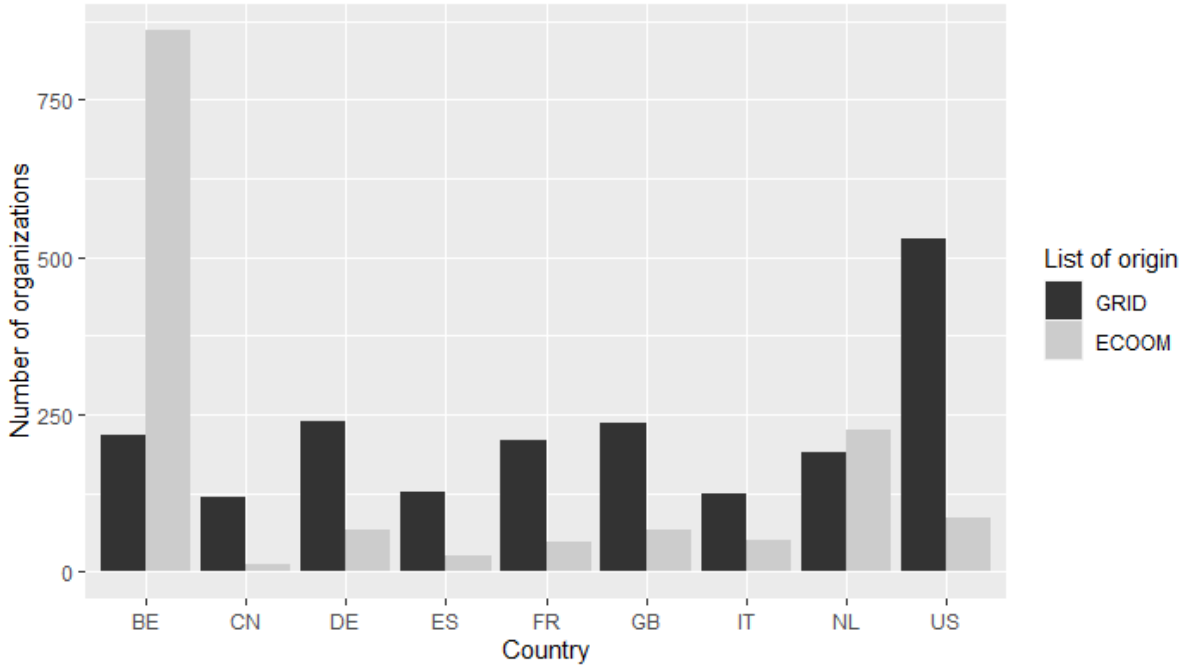
ECOOM, the Centre for R&D Monitoring, administers the *Flemish Academic Bibliography for the Social Sciences and Humanities* (VABB-SHW), a compilation of Flemish SSH research output. In 2019, a new evaluation parameter measuring international collaboration was introduced. Publications of which at least one co-author is affiliated to a foreign institution receive an increased weight in the funding calculation. This parameter is currently based on the Web of Science (WoS) data. The aim is also to collect author affiliation data for all relevant non-WoS VABB-SHW publications for a 10-year timeframe (2010-2019; n=22.121). Although affiliation data for about 10% of these publications can be found in other bibliographic databases, like Crossref or Scopus, the major part needs manual collection efforts.

In 2019, ECOOM launched a data collection operation, using an online application to code the author affiliations. Due to the amount of publications to be processed, affiliations were registered on publication level; in order to guarantee the sustainability and interoperability of the data, the affiliated organizations (universities, NGO's, companies, …) were coded via Digital Science' GRID identifiers. Therefore, the complete GRID database was uploaded to the application. However, as an important part of VABB-SHW publications are non-Anglo-Saxon or locally-oriented, GRID doesn't capture all the affiliated organizations appearing on publications. The online application allows an extension of the initial organization list with new ones, to which also a limited number of GRID variables could be assigned (Type and Country), enabling necessary classification.

At the moment of writing, 90% of the publications are coded. The authors are, in total, affiliated to 5.179 different organizations. However, 1.814 of these organizations were not

included in GRID (35,03%). Almost two thirds of the additional (ECOOM-)organizations are based in Belgium or the Netherlands, which, in absolute numbers, count each more new organizations than GRID-organizations (Figure 1).

Figure 1. Total number of different affiliated organizations by country and list of origin



GRID uses a number of fixed types (Company, Education, Facility, Government, Healthcare, Non-profit, Other) and assigns one to each of the organizations in the list; the additional organizations were coded likewise. GRID covers almost all organizations of the education category (i.e. universities), but lacks data about companies, governmental entities and non-profit organizations (Figure 2). This leaves room for more in-depth analysis, but indicates that the coding of affiliation data for non-WoS publications requires a substantial extension of the organization list. In this case, a GRID+ list is a necessity.

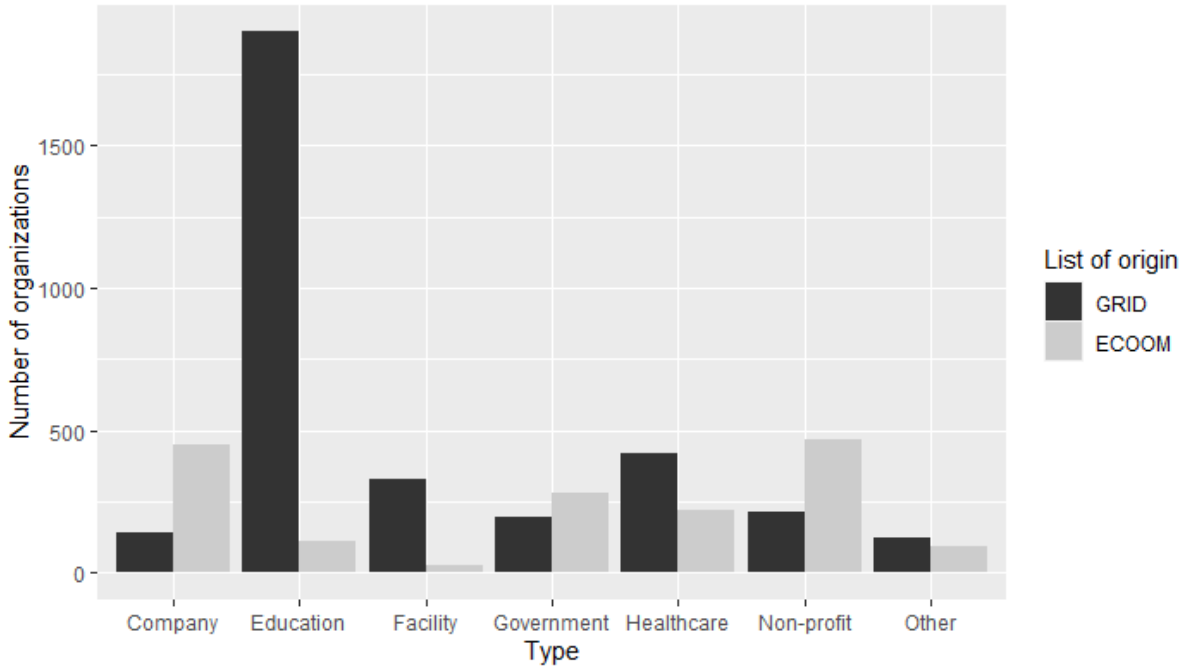
Building on ROR+

Initially, ECOOM's data model was very much aligned with GRID's architecture. This was modified to a more generic model while converting from GRID to ROR. The different dimensions of ROR were kept and extended with modules for adding keywords, registering a chronology or uploading files to an individual record. The labels, aliases and acronyms from ROR were merged into one table. Other types of external identifiers can now be added (e.g. commercial register numbers) as well as additional types/categories; addresses were converted to a 1-n-relationship. It is also being adapted to a specific Belgian context in order to cope with its multilingualism and multiple governmental layers (e.g. the addition of a language variable to acronyms and aliases). Periodic updates from ROR will be integrated in this format. We will remain active in the ROR community and transfer relevant additions or modifications to the register.

The next step consists of a focused deployment and extension of the organization database in different contexts and for different purposes. It could well serve as tool for research projects that make use of (current or historical) organizational data and be enriched by additional organizations and metadata. Administrative services could benefit from ROR-originated intelligence.

It is our aim to present this evolution of an organization database from a tool to code author affiliation data to a customized knowledge instrument on a poster during the conference.

Figure 2. Total number of different affiliated organizations by type and list of origin



Since the end of 2021, GRID is no longer updated for general use. The database was transferred to the community-curated ROR initiative. All GRID identifiers were converted to ROR identifiers in the ECOOM database early 2022. The database continues to evolve as ROR+ ever since.