

# Proof of Concept of a European database for Social Sciences and Humanities publications: The VIRT-ENRESSH pilot

Hanna-Mari Puuska<sup>1</sup>, Tim Engels<sup>2</sup>, Raf Guns<sup>2</sup>, Gunnar Sivertsen<sup>3</sup>, Janne Pölönen<sup>4</sup>, Jorge Mañana-Rodríguez<sup>5</sup>

<sup>1</sup>CSC – IT Center for Science Ltd., Finland

<sup>2</sup>Centre for R&D Monitoring (ECOOM), University of Antwerp, Belgium

<sup>3</sup>Nordic Institute for Studies in Innovation, Research and Education, Norway

<sup>4</sup>Federation of Finnish Learned Societies, Finland

<sup>5</sup>Spanish National Research Council, Spain

## 1. Introduction

This paper reports on a collaborative project in order to investigate a potential cost-efficient solution for the integration of European research information. The project involved partners from Belgium, Finland, Norway, and Spain.

The idea for the project stems from the constant finding in bibliometric research that the most widely used commercial databases, Web of Science or Scopus, do not provide complete coverage of research output in any field, and that particularly in SSH fields they suffer from severe lack of coverage of publications in books and languages other than English (e.g. Sivertsen 2016).

Science policy and research evaluation at all levels of the European Research Area need support from reliable, comparable, and comprehensive information on research activity, productivity and quality. To this end, in the Strategic Membership Meeting of euroCRIS in November 2016, various international organisations (EARMA, Science Europe, OpenAIRE, German Research Academies, JISC, ORCID and euroCRIS) agreed to explore the possibilities to cooperate in order to come to/create a European Research Information Infrastructure.

The issue has been on the agenda in various earlier reports but is still unsolved. An Expert Group on Assessment of University-Based Research recommended in a report to the European Commission that it should “Invest in developing a shared information infrastructure for relevant data to be collected, maintained, analyzed, and disseminated across the European Union” (European Parliament 2010). A report to the European Parliamentary Research Service (2014) recommends development of a European integrated research information system inter-connecting the existing national research information systems.

A recent survey of 41 European countries identified 21 national databases for research output within SSH in 20 countries. The databases differ in terms of their content, openness, and purposes of use (Sile et al. 2017). The main difficulty of standardization and interoperability of data at the European level is the variety of national publication information systems and their data models. Many countries have, however, faced and solved similar problems at national level when they compile information from research organizations using various local systems.

The European Network on Research evaluation in Social Sciences and Humanities ([www.enressh.eu](http://www.enressh.eu)) is an EU funded COST action network with partners from 36 European Countries. The network aims at advancing the understanding of SSH research and facilitate adequate evaluation processes. One of the working groups of ENRESSH is specifically set to coordinate initiatives in terms of standardization and interoperability of

research information in SSH and to design a roadmap for a European database for research outcomes. In view of this task, we have set up a proof of concept of a European database for social sciences and humanities databases. This process, which builds on the efforts made at national level in various countries, is described in this paper. The technical solution of the POC builds on the strengths of the Finnish VIRTAsystem, thus, we refer to this proof of concept as the VIRTAsystem-ENRESSH-POC.

## 2. Origins in Finnish VIRTAsystem publication information service

In Finland, the VIRTAsystem Publication Information Service, launched in 2016, allows Finnish organizations to store a copy of the publication information of their institutional publication databases. The organizations use various local solutions from commercial CRIS systems to self-made publication registers. VIRTAsystem is a data warehouse, "a data hub", making up-to-date metadata from research institutions available for other services and producing comprehensive and comparative information on publishing activity both nationally and institutionally. The technical checking procedure is highly automated including validation of data fields, identification of publication channels and detection of duplicates and co-publications.

The data are used as one criterion of performance-based funding of universities but has also other purposes of use. The publication metadata can be used by research funders, publication or data repositories, infrastructure services, or any other service used by researchers. The major Finnish research funding organization, the Academy of Finland, has introduced import from VIRTAsystem in its project reporting service.

## 3. Implementation of the VIRTAsystem-ENRESSH-POC

The ENRESSH network together with CSC – IT Center for Science in Finland launched the VIRTAsystem-ENRESSH pilot project during the ENRESSH meeting in March 2017 in Sofia, Bulgaria. The target of the pilot was to apply the VIRTAsystem concept for various countries and illustrate its potential for a European Research Information Service. Six universities from four European countries participated in the VIRTAsystem-ENRESSH-POC:

- University of Oslo, Norway
- University of Antwerp, Belgium
- University Carlos III Madrid (UC3M), Spain
- University of Helsinki, Finland
- University of Jyväskylä, Finland
- Tampere University of Technology, Finland

Publication metadata for the years 2014 and 2015 were included in the pilot. Each university submitted either all its publication metadata or the publication metadata from the social sciences and humanities only. Each institution classified all submitted publications into disciplines according to the OECD Fields of Science classification (OECD 2007). Based on a comparison of the data contents in each participating university and country the "lowest common denominator", that is, the data fields that can be found from all participating universities and countries was identified.

So far, there is no shared international standard for publication types. The publication type classifications differ also between the piloting institutions. Quite analogical categories could however be found in all countries. The countries reported their data according to their own classifications and they were mapped into Finnish classification in VIRTAsystem.

## 4. Conclusions and future expansions

A total of over 50.000 references were integrated into the VIRTAsystem-ENRESSH-POC database. In the basis of this data we will highlight the issues encountered during this pilot project, discuss potential improvements

and the possibilities for a structural approach, and illustrate the potential of an integrated European publications infrastructure for science policy at the local, regional, national and European level.

The extension of the concept of Finnish VIRTAs publication information service to other European countries and institutions would provide a potential solution for a European decentralized system aimed at integration and visibility of data about and for the SSH and other fields of science. The current VIRTAs is however heavily based on Finnish data model and some of the data contents are specific to the data collection needs in Finland. By modifying VIRTAs into a more integration based solution and applying international standards (CERIF in particular) it could act as a convenient and cost-efficient way of developing a European publication “data broker”.

An ontology-based data management approach (Daraio et al. 2016) might be implemented across the source databases, hence facilitating future expansion of the dataset into a European research information service. The current dataset already allows to contrast publication patterns in terms of volumes, language use and prestige of publication channels, e.g. between institutions. In the European context with a large linguistic diversity and almost no cross national data integration such possibilities for benchmarking bear huge potential.

The aim is also to link the further development of the VIRTAs project closer with other ongoing integration projects on research evaluation in Nordic countries: 1) the Nordic list project is a cooperative effort in order to operate a register of scholarly and scientific publication channels and 2) biannual bibliometric report evaluating research outcome in Nordic institutions (e. g. NordForsk 2017).

## References

- Daraio, C., Lenzerini, M., Leporelli, C., Naggar, P., Bonaccorsi, A., & Bartolucci, A. (2016). The advantages of an ontology-based data management approach : openness, interoperability and data quality. *Scientometrics*, 108, 441-455, DOI 10.1007/s11192-016-1913-6.
- European Commission (2010). *Assessing Europe’s University-Based Research: Expert Group on Assessment of University-Based Research*. Directorate-General for Research Science in Society 2008 Capacities, 1.4.1. [https://ec.europa.eu/research/science-society/document\\_library/pdf\\_06/assessing-europe-university-based-research\\_en.pdf](https://ec.europa.eu/research/science-society/document_library/pdf_06/assessing-europe-university-based-research_en.pdf)
- European Parliamentary Research Service (2014). *Measuring scientific performance for improved policy making*. Science and Technology Options Assessment. European Parliamentary Research Service. April 2014. PE 527.383. [http://www.europarl.europa.eu/RegData/etudes/etudes/join/2014/527383/IPOL-JOIN\\_ET\(2014\)527383\(SUM01\)\\_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/etudes/join/2014/527383/IPOL-JOIN_ET(2014)527383(SUM01)_EN.pdf)
- NordForsk (2017): *Comparing Research in Social Sciences and the Humanities in the Nordic Countries - an Explorative Study*. [https://www.nordforsk.org/en/publications/publications\\_container/comparing-research-in-social-sciences-and-the-humanities-in-the-nordic-countries-an-explorative-study/view](https://www.nordforsk.org/en/publications/publications_container/comparing-research-in-social-sciences-and-the-humanities-in-the-nordic-countries-an-explorative-study/view)
- OECD (2007): *Revised Fields of Science and Technology Classification in the Frascati Manual*. Paris: Working party of national experts on science.
- Sīle, L., Guns, R., Sivertsen, G., & Engels, T. C. E. (2017). *European Databases and Repositories for Social Sciences and Humanities Research Output*. Antwerp: ECOOM & ENRESSH. Retrieved from <https://doi.org/10.6084/m9.figshare.5172322.v2>

Sivertsen, G. (2016). Patterns of internationalization and criteria for research assessment in the social sciences and humanities, *Scientometrics*, 107:2, 357–368.