

## **PROPOSAL FOR VIRTA EUROPEAN PILOT**

Janne Pölönen (Federation of Finnish Learned Societies), Gunnar Sivertsen (NIFU, Norway) and Hanna-Mari Puuska (CSC – IT Centre for Science Ltd.)

### **ENRESSH WG 3: Databases and uses of data for understanding SSH research**

The European Network for Research Evaluation in the Social Sciences and the Humanities (**enressh.eu**) aims to propose clear best practices in the field of SSH research evaluation. The main objective of working group 3 (ENRESSH WG 3) is to reflect upon the standardisation and the interoperability of current research information systems (CRIS) dedicated to research outcomes from the social sciences and humanities (SSH).

“Confronted by the limitations of the current databases with regard to SSH research (which is circulated in a large variety of languages and often published in other outlets than international peer-reviewed journals), some countries have chosen to develop their own national current research information systems, with complete bibliographic coverage of the scholarly output at research institutions, like CRISTIN in Norway, and VABB-SHW in Flanders.”

“While progress is made in local contexts, there is an obvious need to coordinate the initiatives in terms of standardisation and interoperability of data. Moreover, a European system aimed to a better visualisation of data about and for the SSH remains a necessity.”

“The WG is focused on information systems, databases and repositories for publications and other outcomes of SSH research. The goals of the WG include comparison across SSH disciplines, countries and publication channels, as well as exploration of the feasibility of further streamlining (by developing shared procedures for building and maintaining databases and by designing a roadmap for a European database).”

### **Finnish solution – VIRTA Publication Information Service**

One of the main problems of standardisation and interoperability of data at the European level is the variety of national and institutional CRISes and their data models. This is exactly the situation in Finland, where some higher-education institutions (HEIs) and research institutions use commercial information systems such as Pure and Converis, some have developed their own systems, and others have no CRIS at all.

VIRTA publication information service, launched in 2016, compiles metadata about all publications from all Finnish research organisations. The information is useful in various services and process concerning researchers. VIRTA publication information service will provide the higher education institutions and other research organisations with up-to-date, comparative data on their own publications, as well as those of other institutions. The publication data can be exported to other services or systems via REST API or OAI-PMH from VIRTA, and it is openly available through JUULI-portal ([www.juuli.fi](http://www.juuli.fi)).

HEIs are responsible for data collection to their local CRISes. Data are transferred to VIRTA as XML files once a year or daily, and duplicates are automatically identified. To be admitted from the local CRISes to the VIRTA publication information service, HEIs need to provide for each publication type certain obligatory or voluntary data contents, and the data needs to fulfil certain technical criteria (correct form of ISSN, ISBN,

etc). All the requirements concerning the data have been stated by the Ministry of Education and Culture in the *Publication data collection Guide* to ensure uniformity of data collected from different local CRISes.

### **VIRTA European Pilot**

A possible solution for a European system aimed to integration and visualisation of data about and for the SSH is to extend the Finnish VIRTA publication information service to compile publication data from other European countries and institutions and make them available for various purposes. The technical solutions described above have already been built and developed by CSC – IT Center for Science Ltd., and they are currently operating VIRTA in Finland. There are also data content definitions and data collection instructions needed for standardisation. To copy and modify these technical solutions and definitions would be a convenient and cost-efficient way of developing a European system.

It should be possible to plan, build and have running a pilot of the European Publication Information Service, which could include publication data from national or institutional CRISes of any European country or single organization. This requires simplification and modification of the data content definitions, as some of these are specific to the data collection needs in Finland. Technical solutions to be developed by CSC for the pilot purpose include implementation of a copy of VIRTA, REST API as well as a public web portal. Moreover, an input service can be provided for such organizations that don't have their own CRIS system.

These can be outlined in the proposal, on basis of which it is possible to produce a pilot plan in collaboration between CSC (technical solution) and ENRESSH WG3 (data contents).

Depending on the extent of modifications, the funding of implementing a technical solution for the pilot needs to be arranged. The participating countries or organizations would be responsible for the costs of producing their publication data as XML files and transferring them to VIRTA.

The future maintenance and development of the running service can be funded by a yearly payment for each participating organization or country. Some form of governing body, representing the participating organizations, is needed to steer the operation.

### **Action Points**

Proposal (late 2016 and early 2017)

- Janne, Hanna-Mari and Gunnar (meeting 15.12.2016)

Final Plan (the first half of 2017)

- Data contents - ENRESSH WG 3
- Technical solution - CSC – IT Center for Science Ltd.
- Presentation (Sofia meeting in February, Antwerp conference in July, other?)

Funding

- EU?
- Nordforsk?
- Countries and/or institutions participating in the pilot?

Preparation

- Contents work/Publication data collection Guide - ENRESSH WG 3

- Technical work - CSC – IT Center for Science Ltd.

#### Pilot

- Countries and/or institutions participating in the pilot

#### European Publication Information Service

- Other countries and institutions can join