

Jyrki Ilva: *Building research assessment tools on a national level. The National Library of Finland Bulletin 2011. Helsinki 2011. Permanent address: <http://urn.fi/URN:NBN:fi-fe201108112013>*

Building research assessment tools on a national level

The National Library of Finland has been one of the participating organizations in the planning of a new nationwide current research information system. According to the proposal of the JURE II working group published in June 2011, the new system is to be in use by the year 2014. The system will eventually comprise all of the Finnish universities, including the universities of applied sciences, the state research institutes and the central hospitals.

From numbers to metadata

The main motivation behind the planning of a new current research information system is connected to the research assessment needs of the Ministry of Culture and Education. There are plans to use publication data as one of the main criteria for the funding of Finnish universities and other research organizations. To facilitate this goal, the quality of publication data collected on the national level must be significantly improved.

For many years the universities in Finland, including the universities of applied sciences, have reported the number of research publications by their faculty members to the national databases KOTA and AMKOTA. However, the universities have collected this data using very different tools and methods. To improve the reliability and transparency of the data, it has been decided that, starting from the year 2011, the universities will have to provide the metadata on their publications to the Finnish Ministry of Culture and Education.

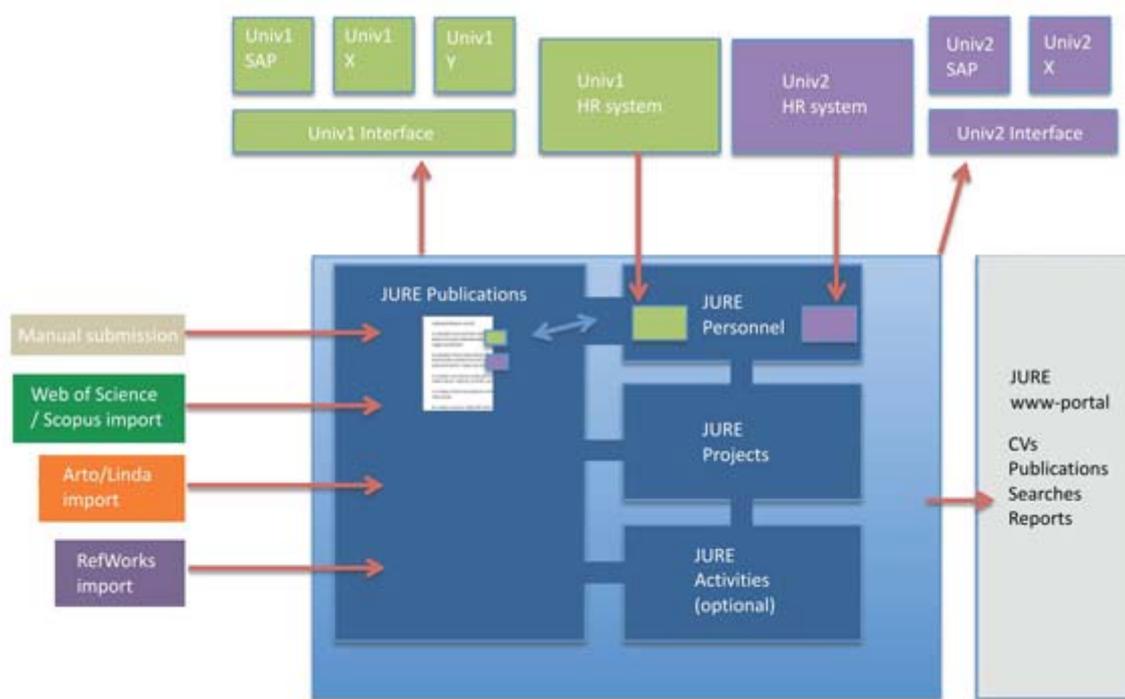
The JURE project (<http://raketti.csc.fi/en/jure>) started in 2009, and its main purpose is the construction of a nationwide publication database. The project is funded by the Ministry of Culture and Education and coordinated by the CSC – IT Center for Science. The CSC is one of the subprojects of the larger RAKETTI project, which seeks to "improve the quality, compatibility, and usability of information and IT solutions in the steering and monitoring of higher education and in the management of higher education institutions". The JURE project is led by a steering group, which includes two representatives from the National Library. In addition, specialists from the National Library have participated in several working groups within the project.

Towards a common national system

Within the JURE project, there has been a great deal of discussion about the construction of the planned national publication database. The key issue has been the degree of centralization needed for a national system. Currently, most Finnish universities have their own local publication databases or research information systems. The question is whether it would be better to build the national system on the existing databases, meaning that the publication data would first be registered in a local database and then collected in a national system, or whether it would be better to create a completely new national system that would replace the existing local solutions. Although setting up of a distributed system based on collecting material from the existing databases would require less work and fewer resources, there are strong incentives for a more centralized model, which seems to be the preferred choice for a long-term solution.

The JURE project is simultaneously working towards two goals. The JURE I working group is planning a short-term publication database for the next few years. This solution will be based on a simple distributed model, which will require only modest investments from the participating organizations. Starting from the year 2011, the Ministry of Culture and Education requires from all Finnish universities publication metadata, not just the number of publications in each of the publication-type categories. Although the normalization of the metadata will require some work, the metadata will be combined with publication metadata harvested from international research databases, and the combined data will be used as one of the starting points in the planning of a new funding model for the universities.

The JURE II working group has been planning a long-term solution for a national current research information system. The final report of the working group was published in June 2011. In the report the working group proposes the building of a centralized nationwide publication database, with optional extra modules for those organizations that would like to gather data on other research activities as well. The proposal includes a preliminary budget, which highlights the benefits of a centralized national system. Although the inclusion of the Finnish universities is considered the first priority, it is likely that the system will be extended to include the scientific output of the universities of applied sciences, the state research institutes and the central hospitals. If the ministries, universities and other organizations agree on the project, then the new system is projected to be in use by the year 2014.



Architecture of the national current research information system (preliminary draft). It will be connected to the information systems of the participating organizations via standardized technical interfaces, and it will also support the harvesting of publication metadata from external sources.

Harvesting metadata from external sources

One of the key requirements for the nationwide current research information system is that it should be capable of ingesting publication metadata harvested from external sources. Currently, most of the publication databases in use at Finnish universities do not support the harvesting of metadata. The most significant exception is Tuhat (<https://tuhat.halvi.helsinki.fi/portal/>), the recently launched new current research information system of the University of Helsinki. Tuhat is based on Pure, a software product developed by Atira (Denmark).

Most of the articles published by Finnish scientists in leading international journals are already being registered in two large commercial databases, the Web of Science (owned by Thomson-Reuters) and Scopus (owned by Elsevier). Each of these databases supports metadata harvesting, although there are contractual restrictions on the use of the data. It would make sense to utilize this metadata in the national current research information system, as it would decrease the amount of local work needed to produce the database and also improve the quality of the data.

FinELib has carried out negotiations with both Thomson-Reuters and Elsevier for the usage rights of the metadata. There have been simultaneous negotiations for two different kinds of deals. On the one hand, the Ministry of Culture and Education wanted to have the full Web of Science raw data for their analyses of longer-term trends in Finnish publication activity. On the other hand, the use of the publishers' metadata in the local organizational databases and in the national database required amendments to existing agreements between the organizations and the publishers.

For contractual reasons, the use of the Web of Science and/or Scopus data involves a definite trade-off: although it offers important benefits, it may also lead to serious restrictions on the use and re-use of publication data.

National databases as sources for Finnish metadata?

International databases such as the Web of Science and Scopus do not cover the national Finnish publication channels. Nevertheless, these channels are very important for some fields of research, and they account for a large portion of Finnish scholarly output as a whole. It has been suggested that the metadata for many of these publications could be harvested from the National Library databases. Norway provides a good illustration of the possibilities of this approach: the number of articles harvested from the Norwegian national article database accounts for 25 per cent of all articles in the Norwegian national publication database.

The National Library of Finland has investigated the possibilities of using the national databases Linda and Arto as sources for the metadata. Linda is the union database of all Finnish university libraries, while Arto is a national article database. According to the report of this investigation, the databases could be used as metadata sources, but there are also challenges and developmental needs in such an approach.

The obvious challenges are the scope and timeliness of the metadata production. Arto is especially illuminating in this respect. The cataloguing of scholarly publications in Arto is currently done on a volunteer basis by the participating libraries, each of which has pledged to catalogue the contents of certain journals. An obvious problem is that there are many journals that currently are not being catalogued. Another challenge is that the Arto work is usually not the primary duty of any single librarian, and other duties are often deemed more urgent. Consequently there are often delays (of weeks or months) before the articles in a new issue of a journal are catalogued in Arto. From the point of view of a current research information system, it is important to obtain the metadata of new articles as quickly as possible, because otherwise the information may have to be added manually, leading to unnecessary duplication of work.

Currently, databases such as Arto and Linda are mainly intended for libraries and information seekers. The records in these databases lack clearly defined affiliation and publication-type data, which are essential for research assessment and bibliometric purposes. Adding this data would mean extra work for the cataloguers, even if such information were available in the publication itself (which is not always the case).

The investigation made by the National Library of Finland suggested that co-operation with Finnish scholarly publishers would help to solve these problems. The ideal solution would be to get as much publication metadata as possible directly from the publishers and journal editors. Obviously, this is not a simple task. Compared to major international publishers, the Finnish scholarly publishers are tiny. Most of them publish

only one or two journals, and they operate on a more or less limited budget. Few of them have the kind of information system that could be used for metadata harvesting.

Although the publishers or journal editors would have to do extra work to provide the metadata for their articles, it seems fairly likely that many, if not most of them, would find it worthwhile, since to do so significantly increases the visibility of their publications among the key audiences. Of course, some technical development is also needed to make this possible. Arto is currently a Voyager database, and since the Voyager cataloguing client is not a suitable option for non-librarians, a new ingest system that works outside Voyager is required. The ingest system should be designed to be as easy to use and highly optimized for the task as possible.

Rating the channels of publication

The Ministry of Education and Culture has also funded the Publication Forum project, which is coordinated by the Federation of Learned Societies. The project will produce ratings of 22,000 scholarly journals, publication series and publishers by the end of the year 2011, following the model in use in several other countries, including Norway and Denmark. In practice this work is being done by 23 panels, the members of which are leading scientists, mostly university professors, in each field. The project has a close connection to JURE, as the ratings produced by the project will also be used in the national publication database.

The aim of the Publication Forum is to identify the scientific publishing channels used by Finnish scholars and classify them into two or three categories. The first category (level 1) consists of channels recognized as scientific. The second category (level 2) contains the leading publication channels in each field. It has been agreed that this group may not contain more than 20 per cent of all scientific publication channels in each field. In addition, there has also been discussion of whether there is a need for a level 3, which would contain a small number of truly outstanding publication channels. Some of the panels have been less than enthusiastic about this idea, and the decision on the use of this level has been left to each panel.

Not unexpectedly, one of the issues that has generated discussion is the treatment of Finnish publication channels in the ratings. One of the aims of the whole rating exercise is to boost the level of ambition among Finnish scholars by making it more rewarding for them to publish their research findings in top international journals. However, in some fields a large proportion of the entire scholarly output is published through Finnish publication channels, and it may be very tempting for the panelists to take this to account in their ratings. Of course, in some fields such as Finnish language and Finnish history, it can be argued that some of the Finnish journals or publishers are among the leading publication channels, even from an international perspective, as they reach most of the members of the research community active in these fields.

Jyrki Ilva is an Information Systems Specialist at the National Library of Finland.