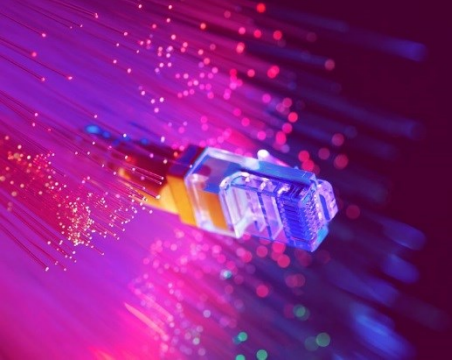




Proof of Concept of a European database for Social Sciences and Humanities publications:

The VIRT-A-ENRESS-POC



Hanna-Mari Puuska, CSC – IT Center for Science, Finland

Tim Engels, University of Antwerp, Belgium

Raf Guns, University of Antwerp, Belgium

Janne Pölönen, Federation of Finnish Learned Societies

Gunnar Sivertsen, NIFU, Norway

Jorge Mañana-Rodriguez, Spanish National Research Council



CSC – Finnish research, education and public administration ICT knowledge centre

The framework and set up of the project: ENRESSH network



- The European Network on Research evaluation in Social Sciences and Humanities (www.enressh.eu) is an EU funded COST action network with partners from 36 European Countries
- ENRESSH aims to propose best practices in the field of SSH research evaluation
- One of the goals of the ENRESSH is to design a roadmap for a European database for SSH output
 - In view of this task, a proof of concept VIRT-A-ENRESSH-POC of a European database for publications was set up

VIRTA-ENRESSH-POC



- VIRTA-ENRESSH-POC is a collaborative pilot project exploring a potential cost-efficient solution for the integration of European research information
 - Especially for SSH but not excluding other fields
 - Carried out between 3/2017-3/2018
 - Involved partners from Belgium, Finland, Norway, and Spain
 - Founded on the efforts made at national level in participating countries
 - The technical solution builds on the strengths of the Finnish VIRTA Publication Information Service

Challenges of integration

- 21 national databases for research output within SSH in Europe (*Sile et al. (2017). European Databases and Repositories for Social Sciences and Humanities Research Output. Antwerp: ECOOM & ENRESSH.*)
- The national databases differ in terms of their content, openness, and purposes of use
- The main difficulty of standardization and interoperability of research information at the European level is the variety of national systems, processes and data models.
- Many countries are facing a similar problem at national level when they compile information from research organizations using various local systems, e.g.:
 - Norway: a national CRIS used by all organizations
 - Flanders & Finland: data integrated from various local CRISes to a regional database

VIRTA - Finnish solution for integrating publication metadata



- The publication metadata are compiled in **VIRTA Publication information Service** where 54 Finnish organizations exchange a copy of all publication information in their institutional CRISes
- VIRTA is a data warehouse, “a data hub”, making publication information available for other services and providing up-to-date, comprehensive and comparative data on publishing activity nationally and institutionally
 - Publication information available for automatized imports to research funding reports of funders’ services etc.

Finnish VIRT A Publication Information Service – key features



Data sources	Local CRISes or publication databases of HEIs, university hospitals, state research institutes
Data format	XML files (CSV converter and an input service provided for small organizations)
Data contents	The data must include required fields and fulfil certain technical criteria defined in <i>VIRT A XML schema</i> .
Data validation	Duplicates and co-publications, missing fields and errors as well as publication forums identified automatically and real time. Error reports available for research organizations in an online service.
Data transfer	From organizations via a secure and certified connection by using SFTP protocol and SSH authentication keys.
Updates	New publications and corrections in local systems updated automatically to VIRT A. The frequency depends on the organizations. All data from previous years to present can be transferred.
Data use and availability	Metadata exported daily to www.juuli.fi . Statistics compiled once a year in www.vipunen.fi APIs: REST and OAI-PMH (CERIF API on development)

VIRTA-ENRESSH-POC– practical issues



Participating countries reported their complete publication metadata from the years 2014-15:

- **Norway:** University of Oslo
 - **Flanders:** University of Antwerpen
 - **Spain:** Universidad Carlos III de Madrid (UC3M)
 - **Finland:** University of Helsinki, University of Jyväskylä, Tampere University of Technology
-
- 52 948 publications in total
 - Finland and Oslo cover all fields, Antwerpen and Madrid only SSH

Data format:

- The pilots exported their own data into a CSV model file and converted the file into VIRTA XML schema by using a CSV-XML tool.
- Only the core information were required as mandatory: publication title, publication year, authors, publication type, field of science, organization authors (other fields were optional)

Issues of data comparability:

1. Disciplines



- The pilots mapped their publications into OECD Frascati Manual's FoS classification
- A mapping procedure was quite easy to apply but there is variation in the definitions of the fields, being determined by
 1. publication itself
 2. the journal of the publication
 3. the author of the publication
 4. the organizational unit of the author

Issues of data comparability:

1. Inclusion criteria, semantics and publication types

- The countries vary in terms of their inclusion criteria, e.g.
 - Scientific only or non-scholarly publications as well (professional and popular books, articles, reports etc.)?
 - Conference presentations, short abstracts included?
- A mapping procedure for publication types can be applied but still the data are not fully comparable since the definitions of for example "article", "book chapter" or "scientific" vary
- Conclusion:
 - Agreement on semantics and publication types amongst all countries probably not feasible
 - Authorized publication channel registries as a solution for more structured and comparable data?

Issues of data comparability: Publication type mapping

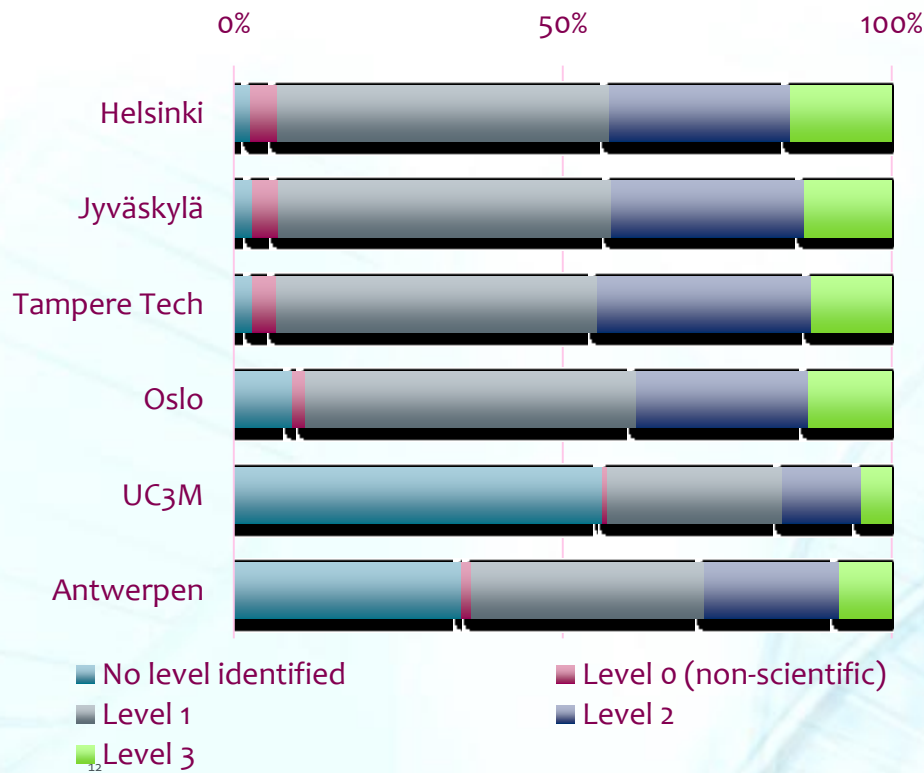
		Flanders 1=peer-reviewed / 0 = non peer-reviewed		Norway
Finland / Madrid				
Peer-reviewed articles	A1 Journal article, original research	VABB-1: journal article	1	3= Article in series (ISSN)
	A2 Review article			
	A3 Book section	VABB-4: book chapter	1	2= Article in book (no ISSN)
	A4 Conference proceedings	VABB-5: proceedings paper	1	
Non peer-reviewed articles	B1 Non-refereed journal articles	VABB-1: journal article	0	
	B2 Book section	VABB-4: book chapter	0	
	B3 Non-refereed conference proceedings	VABB-5: proceedings paper	0	
Monographs	C1 Book	VABB-2: monograph	1	1= Monograph
	C2 Edited book	VABB-3: edited book	1	
Professional	D1 Article in a trade journal			
	D2 Article in a professional book			
	D3 Professional conference proceedings			
	D4 Development or research report			
	D5 Textbook, professional manual or guide			
	D6 Edited professional book			
Popular	E1 Popularised article, newspaper article			
	E2 Popularised monograph	VABB-2: monograph	0	
	E3 Edited popular book	VABB-3: edited book	0	

Authorized publication channel registries as a solution for more structured and comparable data?

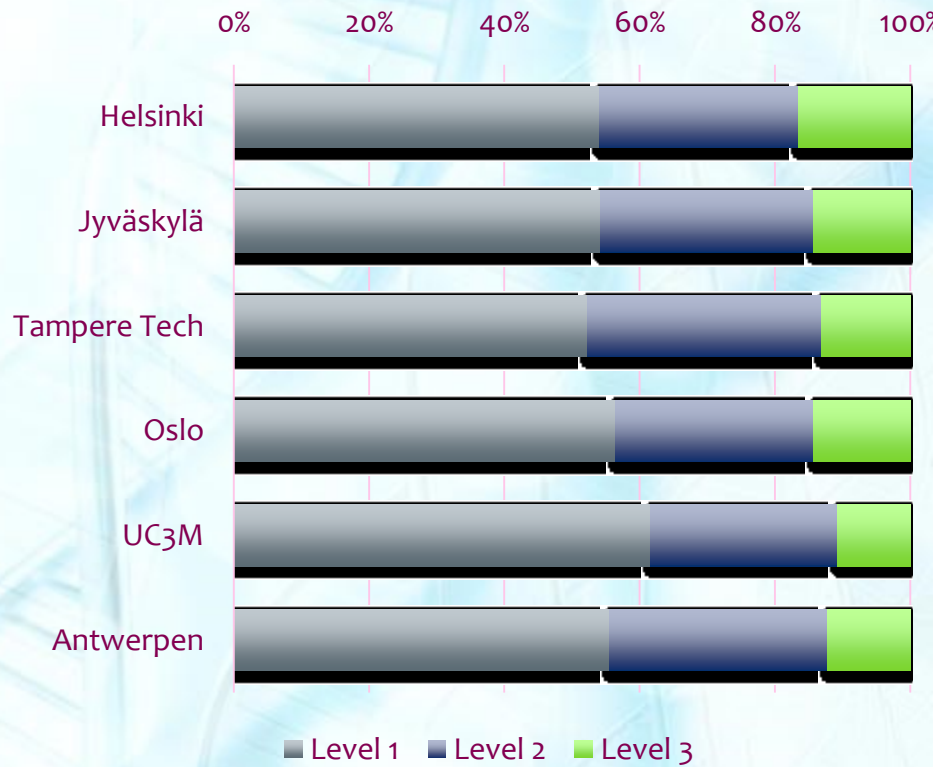
- The data collected in the pilot has its highest quality and consistency in terms of the bibliographic data meanwhile the classifications vary
- For all publications reported in the POC, the publication channel was automatically detected against the Finnish Publication Forum database (JUFO)
 - JUFO
 - used for publication channel rankings as part of universities' funding model
 - is integrated with other relevant databases (e.g. ISSN, DOAJ and ERIH) and
 - contains structural data on journals and series, conference proceedings and book publishers
 - includes information on type (scholarly/non-scholarly), open access policy, peer-review practice, scientific fields and internationality etc...
- Corresponding registries for publication channels are maintained also in other countries, such as Norway, Denmark and Belgium (Flanders).
- “The Nordic List” funded by NordForsk has implemented a common Nordic registry of



Journal articles, all



Journal articles, identified as level 1-3



Next steps



- Collaboration to be continued both in the framework of 1) ENRESSH and 2) Nordic countries
- In a Nordic meeting in Finland in May 2018, the stakeholders of national CRIS systems in Nordic countries decided to
 - continue both contextual and technical development of ENRESSH-VIRTA
 - integrate it with other ongoing NordForsk's integration projects on research information management: 1) the Nordic list and 2) bibliometric analysis comparing Nordic institutions in SSH fields.
- The cooperation at Nordic level does not exclude other European countries and the next goal is also to extend the POC to more countries.
- The next phase also includes investigation of the use of CERIF, in import and export in ENRESSH-VIRTA.
- Cooperation to be strengthened also with other initiatives that aim at the integration of publication metadata at European level



Hanna-Mari Puuska

Development Manager, PhD

hanna-mari.puuska@csc.fi

+358 50 3818 568



<https://www.facebook.com/CSCfi>



<https://twitter.com/CSCfi>



<https://www.youtube.com/c/CSCfi>



<https://www.linkedin.com/company/csc---it-center-for-science>