

Integration of national publication databases Towards a high-quality and comprehensive information base on scholarly publications in Europe

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

Joonas Nikkanen, CSC – IT Center for Science, Finland

Tim Engels, University of Antwerp, Belgium

Raf Guns, University of Antwerp, Belgium

Dragan Ivanović, University of Novi Sad, the Republic of Serbia

Janne Pölonen, Federation of Finnish Learned Societies, Finland



Integration of national publication databases –

Towards a high-quality and comprehensive information base on scholarly publications in Europe

Contents

1. European scholarly publication infrastructure – why?
2. Potential of the existing bibliographic databases in research evaluation of SSH
3. National publication databases & publication channel databases
4. VIRTAs-ENRESSH Proof of Concept for a European scholarly publication infrastructure

European scholarly publication infrastructure – why?

- “ Science policy and research evaluation of the **European Research Area** need reliable, and comparable information on research activities, productivity and quality.
- “ The call for complete **open access** in EU member states and the **Plan S** initiative highlight the need for reliable information to follow up on the development of open access publishing.
- “ **DORA - Declaration on Research Assessment, the Leiden Manifesto** for research metrics and **the Metric Tide report** recommend that the research evaluation should be based primarily on expert judgment, but may be supported by responsible metrics which presume *robustness, transparency, diversity, and reflectivity of the data*.



However, comprehensive information on European publishing activities is still missing...

Potential of the existing bibliographic databases in monitoring and assessing research in SSH



- **Web of Science** and **Scopus** citation index databases
 - + quality-controlled and reliable data, strong focus in English language scientific journals
 - lack of coverage especially in SSH, namely books and national publications
- The aggregating harvesters, e.g. **Google Scholar**, **Microsoft Academic**, and **OpenAIRE**
 - + comprehensive in terms of scientific fields
 - coverage is rather 'accidental' than systematic
 - GS & Academic do not disclose their sources -> it is not known which publications are left out
 - not possible to assess to what extent biased in terms of languages, publication types, open access, or scientific fields
- Services relying on DOI, e.g. **Dimensions** and **Unpaywall**
 - + coverage not limited to international journals only
 - considerable differences in DOI availability between different publication types, fields, and countries

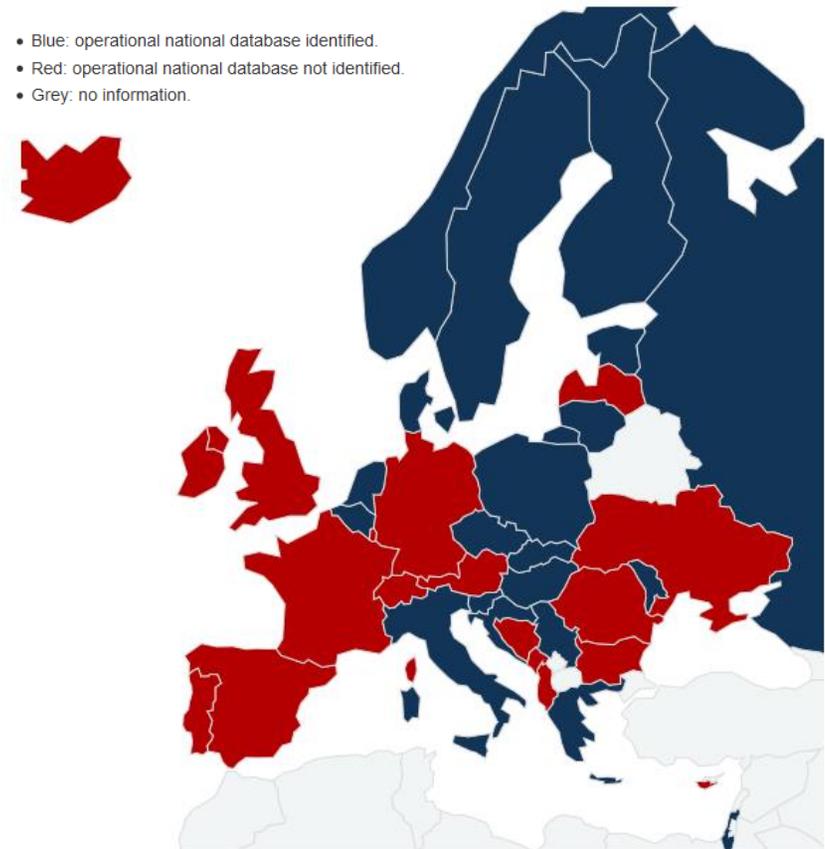
So far, the existing databases do not meet the needs of research evaluation especially in SSH fields.

National bibliographic databases

- “ Since the other databases do not give a comprehensive picture of publication activities, European countries have invested in national research information infrastructures.
- “ The aim is to better monitor and assess the research performance at national level.
- “ A survey by Sile et al.¹ found out that **20 European countries** have a national database for research output. Many of them cover all fields of science.
- “ The national databases are either
 1. Centralized information systems into which all organizations input metadata on publications and other research output, or
 2. Harvest information from research organizations’ own CRISes (Current Research Information Systems).
- “ Many countries use their national publication database as a basis for universities’ governmental performance-based research funding allocation.
- “ National databases often have a mandatory nature which ensures their comprehensiveness, quality assurance, and precise documentation.

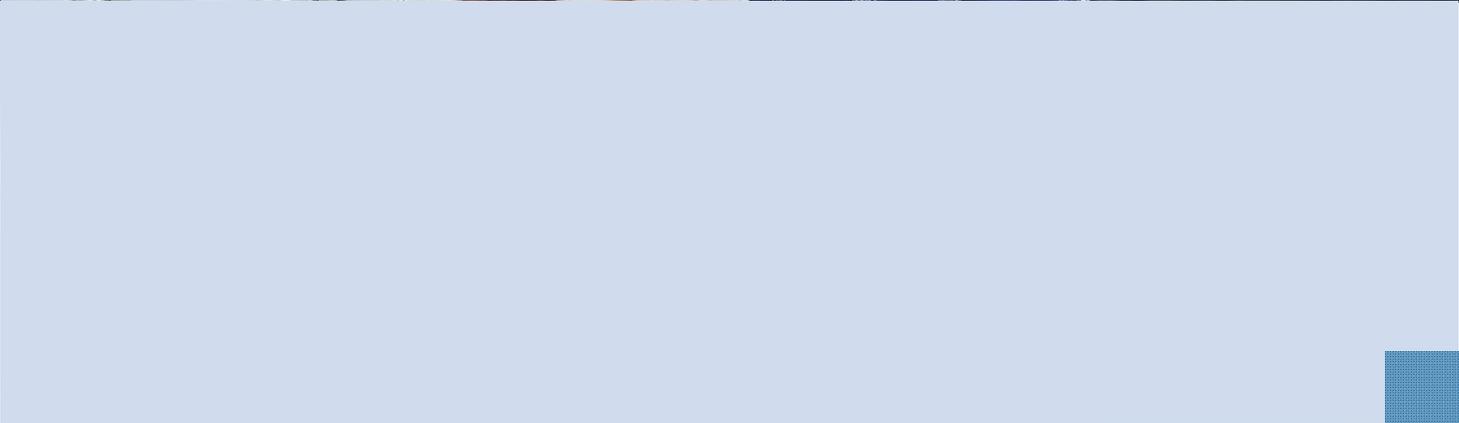
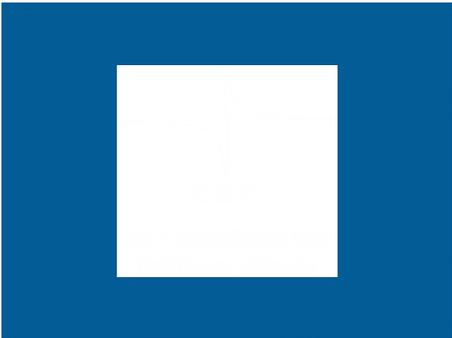
¹Sile, L., Guns, R., Sivertsen, G., & Engels, T. C. E. *European Databases and Repositories for Social Sciences and Humanities Research Output*. Antwerp: ECOOM & ENRESSH (2017). <https://doi.org/10.6084/m9.figshare.5172322.v2>

- Blue: operational national database identified.
- Red: operational national database not identified.
- Grey: no information.



Source:

<https://ecoom.uantwerpen.be/sites/en/edrssh/o/europeandatabasesmap>



Publication channel databases

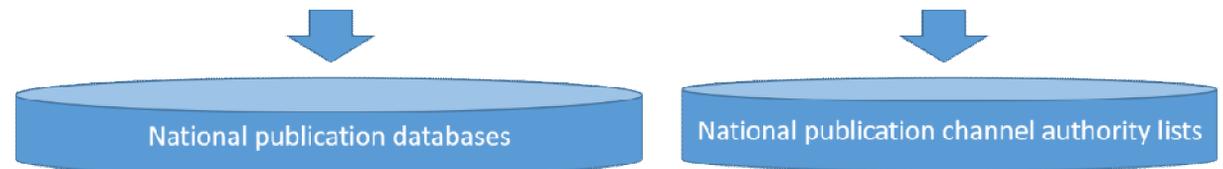
- “ In many European countries, the national publication databases are accompanied by comprehensive authority lists of **publication channels** = journals, series, conferences, and publishers.
- “ These databases usually include rankings of publication channels for national research assessment or funding allocation purposes.
- “ Often integrated to other databases such as ISSN, DOAJ, ERIH...
- “ The publication channel databases can be used in determining publication types, peer-review status, open access status, and scientific fields of the publications, and thereby, they help to ensure consistency of national publication databases.

Publication metadata

Publication title	Publication patterns in the social sciences and humanities: evidence from eight European countries
Publication year	2018
Authors	Emanuel Kulczycki, Tim C. E. Engels, Janne Pölonen, Kasper Druun, Marta Dušková, Raf Guns, Robert Nowotriak, Michal Petr, Gunnar Sivertsen, Andreja Isterič Starčič, Alesia Zuccala
Organizations	Mickiewicz University in Poznań; University of Antwerp; Federation of Finnish Learned Societies; Danish Ministry of Higher Education and Science; Slovak Centre of Scientific and Technical Information; Lodz University of Technology; Masaryk University; Nordic Institute for Studies in Innovation; University of Ljubljana; University of Primorska; University of Copenhagen
Keywords	Publication patterns, Social sciences, Humanities, Language, Publication type, Internationalization
Journal name	Scientometrics
ISSN	0138-9130
Volume	116
Issue	1
Pages	463-486
Language	English
Publication type	Article in scientific journal
Peer-reviewed	Yes
Open access status	Gold

Publication channel metadata

ID	66909
Title	Scientometrics
Type	Serial
ISSN (print)	0138-9130
ISSN (online)	1588-2861
Start year	1978
Country of publication	Hungary
Publisher	Springer
Language	English
Norway level	1
Denmark level	2
Finland level	2
DOAJ	No
Sherpa/Romeo	Green
Scientific fields	113 Computer and information sciences 513 Law 518 Media and communications 520 Other social sciences



The publication channel databases should also be integrated at the European level!

VIRTA-ENRESSH PoC for a European scholarly publication infrastructure



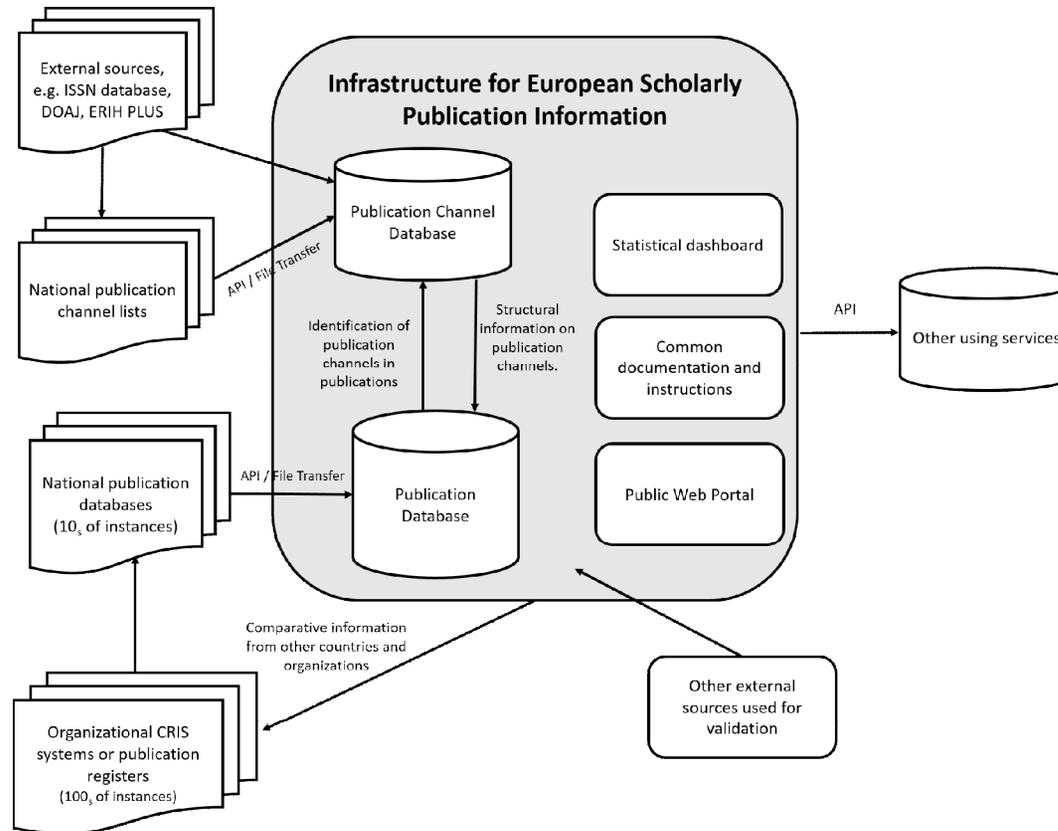
VIRTA-ENRESSH¹ Proof of Concept



- Aim: To demonstrate the integration of publication data from different countries
- Compiled publication metadata from 6 universities from Belgium, Finland, Norway, and Spain
- Especially for SSH but other fields were not excluded
- Founded on the efforts made at national level in participating countries
- Carried out between 3/2017-6/2019
- Results:
 - A Proof of Concept database of European scholarly publications
 - A technical solution and data transfer process built on the Finnish “VIRTA” system
 - In Finland, the VIRTA Publication Information Service compiles information from different organizational CRIS systems and publication registers
 - A minimum metadata model - “the lowest common denominator” – compliant with CERIF (the *Common European Research Information Format*)

¹ ENRESSH (The European Network on Research evaluation in Social Sciences and Humanities, www.enressh.eu) was an EU funded COST action network with partners from 36 European Countries

Integration of national publication databases – solution from VIRTA-ENRESSH-PoC



Results from VIRTA-ENRESSH Proof of Concept

- “ The Finnish VIRTA concept works for cross-national integration of publication databases
- “ The core bibliographic data is consistent across countries.
- “ Variation, however, in
 - “ The extent of other information
 - “ Definitions of e.g. open access, publication types, and scientific fields
- “ A solution based on agreement and shared typologies at European level is not realistic (since often linked to a national framework or legislation)
- “ Instead, additional information should be determined by using other sources such as the authorized publication channel lists
- “ Also, other external data sources can be used to check, e.g. the open access status of publications
- “ Machine learning algorithms can be used to identify the subject fields of the publications.

	Finland / Madrid	Flanders 1=peer-reviewed / 0 = non peer-reviewed	Norway
Peer-reviewed articles	A1 Journal article, original research	VABB-1: journal article	3= Article in series (ISSN)
	A2 Review article		
	A3 Book section	VABB-4: book chapter	2= Article in book (no ISSN)
	A4 Conference proceedings	VABB-5: proceedings paper	
Non peer-reviewed articles	B1 Non-refereed journal articles	VABB-1: journal article	
	B2 Book section	VABB-4: book chapter	
	B3 Non-refereed conference proceedings	VABB-5: proceedings paper	
Monographs	C1 Book	VABB-2: monograph	1= Monograph
	C2 Edited book	VABB-3: edited book	
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

No need for unifying national field classifications!



More information:

Puuska, H. M., Guns, R., Pölönen, J., Sivertsen, G., Mañana-Rodríguez, J., & Engels, T. *Proof of concept of a European database for social sciences and humanities publications: description of the VIRTAs-ENRESSH pilot*. ENRESSH report (2018). <https://doi.org/10.6084/m9.figshare.5993506>

Thank you!



Contact: hanna-mari.puuska@csc.fi



facebook.com/CSCfi



twitter.com/CSCfi



youtube.com/CSCfi



linkedin.com/company/csc--it-center-for-science



github.com/CSCfi