

# overview of the ENRESSH good practices manual for national bibliographic databases



Linda Sīle, doctoral researcher

Centre for R&D Monitoring (ECOOM), University of Antwerp, Belgium

ENRESSH Training school : Poznan, 22/10/2019



#### CREATING AND MAINTAINING A NATIONAL BIBLIOGRAPHIC DATABASE FOR RESEARCH OUTPUT

Result of collaborative work between:

Linda Sīle Centre for R&D Monitoring (ECOOM), University of Antwerp, Belgium

Raf Guns Centre for R&D Monitoring (ECOOM), University of Antwerp, Belgium

Dragan Ivanović University of Novi Sad, Serbia

Janne Pölönen Federation of Finnish Learned Societies, Finland

Tim C.E. Engels Centre for R&D Monitoring (ECOOM), University of Antwerp, Belgium

Manual of good practices ENRESSH & ECOOM



# Outline

#### Background

- -- good practices in research evaluation
- -- ENRESSH surveys, ENRESSH-VIRTA pilot, workshop
- -- goals and scope
- -- challenges and limitations

Content

- -- overview
- -- purpose
- -- stakeholders
- -- research output types
- -- adapting to changes

# Background

#### Good practices in research evaluation

#### Leiden Manifesto

"construction of the databases required for evaluation should follow clearly stated rules, set before the research has been completed"

"To ensure data quality, all researchers included in bibliometric studies should be able to check that their outputs have been correctly identified. Everyone directing and managing evaluation processes should assure data accuracy, through self-verification or third-party audit"

#### San Francisco Declaration on Research Assessment (DORA)

- "II. Be open and transparent by providing data and methods used to calculate all metrics"
- "12. Provide the data under a licence that allows unrestricted reuse, and provide computational access to data, where possible"
  - "14. Account for the variation in article types (e.g., reviews versus research articles), and in different subject areas when metrics are used, aggregated, or compared"

#### Two ENRESSH surveys on databases

- Survey #1 identification of national bibliographic databases for social sciences and humanities scope: 41 countries 95% response rate
- Survey #2 a detailed description of comprehensiveness and data processing scope: 17 databases 76% response rate



#### European databases and repositories for Social Sciences and Humanities research output

Report July 201



#### Comprehensiveness of national bibliographic databases for social sciences and humanities: Findings from a European survey

Linda Sile<sup>1,\*</sup>, Janne Pölönen<sup>2</sup>, Gunnar Sivertsen<sup>3</sup>, Raf Guns<sup>1</sup>, Tim C. E. Engels<sup>1</sup>, Pavel Arefiev<sup>4</sup>, Marta Dusková<sup>5</sup>, Lotte Faurbæk András Holl<sup>7</sup>, Emanuel Kulczycki<sup>8</sup>, Bojan Macan<sup>9</sup>, Gustaf Nelhans<sup>10</sup> Michal Petr<sup>11</sup>, Marieta Pisk<sup>12</sup>, Sándor Soós<sup>7</sup>, Jadranka Stojanovski<sup>9,13</sup> Ari Stone<sup>14</sup>, Jaroslav Šušol<sup>15</sup> and Ruth Teitelbaum<sup>16</sup>

Research Evaluation, 2018, 1-12 doi: 10.1093/reseval/nvv016 Article

<sup>1</sup>Centre for R&D Monitoring (EC00M), Faculty of Social Sciences, University of Antwerp, Antwerp 2020, Belgium, <sup>3</sup>Federation of Finnish Learned Societies, Helsinki 00170, Finland, <sup>3</sup>Nordic Institute for Studies in Innovation, Research and Education, Oslo 0608, Norway, <sup>4</sup>Scientific Electronic Library—eLIBRARY.RU, Moscow 117246, Russia, <sup>8</sup>Slovak Centre of Scientific and Technical Information, Bratislava 811 04, Slovak Republic, <sup>10</sup>Danish Agency for Science and Higher Education, Copenhagen K DK-1260, Denmark, <sup>7</sup>Department of Science Policy and Scientometrics, Library and Information Centre of the Hungarian Academy of Sciences (MTA), Budapest 1051, Hungary, \*Scholarly Communication Research Group, Faculty of Social Sciences, Adam Micklewicz University in Poznań, Poznań 61-712, Poland, <sup>®</sup>Centre for Scientific Information, Rudjer Boskovic Institute, Zagreb 10000, Croatia, <sup>10</sup>Swedish School of Library and Information Science, University of Borås, Borås S-501 90, Sweden, <sup>11</sup>Research Office, Masaryk University, Bmo 601 77, Czech Republic, <sup>13</sup>Research Centre of the Slovenian Academy of Sciences and Arts, Ljubijana 1000, Slovenia, <sup>13</sup>Department of Information Sciences, University of Zadar, Zadar 23000, Croatia, <sup>14</sup>The Planning and Budgeting Committee, The Council for Higher Education, Jerusalem 91000, Israel, <sup>16</sup>Faculty of Arts, Comenius University in Bratislava, Bratislava 814 99, Slovak Republic and <sup>16</sup>The Henrietta Szold Institute, Jerusalem 91000, Israel

\*Corresponding author. Email: Linda.Sile@uantwerpen.be

Abstract

This article provides an overview of national bibliographic databases that include data on research output within social sciences and humanities (SSH) in Europe. We focus on the comprenensiveness of the database content. Compared to the data from commercial databases such as Web of Science and Scopus, data from national bibliographic databases (e.g. Flemish Academic Bibliographic Database for the SSH (VABB-SHW) in Belgium, Current Research Information System in Norway (CRISTIN)) are more comprehensive and may, therefore, be better fit for bibliometric analyses. Acknowledging this, several countries within Europe maintain national bibliographic databases; detailed and comparative information about their content, however, has been limited. In autumn 2016, we launched a survey to acquire an overview of national bibliographic databases for SSH in Europe and Israel. Surveying 41 countries (responses received from 39 countries), we identified 21 national bibliographic databases for SSH. Further, we acquired a more detailed description of 13 databases, with a focus on their comprehensiveness. Findings indicate that even though the content of national bibliographic databases is diverse, it is possible to delineate a subset that is similar across databases. At the same time, it is apparent that differences in national bibliographic databases are often bound to differences in country-specific arrange ments. Considering this, we highlight implications to bibliometric analyses based on data from

CThe Authorial 2018. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Oxation Commons Attribution License (http://oranive.ommons.org/Seansec/by(4.0), which pamits unserticide tense, distribution, and representation in any motion in, provided the original work is properly cited.

Sile, L., Guns, R., Sivertsen, G., & Engels, T. C. E. (2017). European Databases and Repositories for Social Sciences and Humanities Research Output (p. 25). Retrieved from ECOOM & ENRESSH website: https://doi.org/10.6084/m9.figshare.5172322.v2 Sīle, L., Pölönen, J., Sivertsen, G., Guns, R., Engels, T. C. E., Arefiev, P., ... Teitelbaum, R. (2018). Comprehensiveness of national bibliographic databases for

social sciences and humanities: Findings from a European survey. Research Evaluation, 27(4), 310–322. https://doi.org/10.1093/reseval/rvy016

# **ENRESSH-VIRTA** Proof of Concept



Proof of concept of a European database for social sciences and humanities publications: Description of the VIRTA-ENRESSH pilot

ENRESSH

Report March 2018 a collaborative pilot project exploring a potential cost-efficient solution for the integration of European research information for SSH but not excluding other fields carried out between 3/2017-3/2018 with partners from Belgium, Finland, Norway, and Spain

builds on the strengths of the Finnish VIRTA Publication Information Service

Puuska, H.-M., Guns, R., Pölönen, J., Sivertsen, G., Mañana-Rodríguez, J., & Engels, T. (2018). *Proof of* concept of a European database for social sciences and humanities publications: Description of the VIRTA-ENRESSH pilot (p. 23). Retrieved from CSC & ENRESSH website: https://doi.org/10.6084/M9.FIGSHARE.5993506

# Workshop in Antwerp, Belgium

#### Working with national bibliographic databases for research output

#### 10 - 11 September 2018

#### What is this about?

The main purpose of this workshop is to bring together people who are directly involved in the maintenance of national bibliographic databases and exchange experience on various aspects of this work. Special focus of this workshop is challenges (and ways to address them) in data collection and processing:

- How scholarly publications are identified?
- How data from different sources of data are integrated?
- How the accuracy of metadata for records on publications is checked?
- How collaboration with governmental bodies and/or authors of research output is organised?

The programme of the workshop entails presentations addressing several aspects of national bibliographic databases as well as two experience exchange sessions devoted, specifically, to data collection and processing.

We hope this workshop will, first and foremost, serve as a networking site for operators of national bibliographic databases for research output. In broader terms, this workshop aims to facilitate identification of a variety of practices that can be used to address challenges which inevitably are present when operating a nation-wide database.

We welcome participants willing to actively engage into an exchange of experience with or without a presentation on the database they are working with.

Further questions? Contact Linda Sile: linda.sile[at]uantwerpen.be





Sīle, L., Guns, R., & Engels, T. C. E. (2018, November 13). Towards more consistent, transparent, and multi-purpose national bibliographic databases for research output. Retrieved 14 February 2019, from LSE Impact Blog website: <u>https://blogs.lse.ac.uk/impactofsocialsciences/2018/11/13/towards-</u>more-consistent-transparent-and-multi-purpose-national-bibliographic-databases-for-research-output/

## The concept of the manual

Manual of best practices Manual of GOOD practices

30 issues to think about in database design, organisation, maintenance, and usage

! For research evaluation! For social sciences and humanities

A source of inspiration, a trigger for a discussion NOT a step-by-step guide

## Challenges and limitations

Challenge #1 : multiple standards in terminology

One and the same aspect can be named and conceptualised in multiple ways

Metadata schema or data model?

Stakeholder engagement or service design?

Semantic interoperability or cross-cultural validity?

Challenge #2 : choosing the right level of detail

Different user groups of the manual require different level of detail *in different* directions

More technicalities

More theory from librarianship

More on legal issues



## Overview



I. Identify and make explicit the purpose(s) of the database
7. Collaborate with stakeholders
10. Aim for inclusion of a wide range of research output types
30. Follow and adapt to developments in research practices, research policy, and database maintenance

#### I. Identify and make explicit the purpose(s) of the database

#### Helps to design the database in line with users' needs



http://s3.amazonaws.com/stripgenerator/strip/23/50/96/00/00/full.png

#### Different emphases in purpose (database logics)

Enlightenment	"Furthermore, there is another group of beneficiaries - located between the scientists and the layman citizen - these are the teachers, students, hobby scientists. <u>The portal</u> <u>functions of the [database] can lead them to scientific results</u> - publications residing in repositories - they can understand and use in their studies."
New Public Management	<u>"The most important criteria for project evaluation and subsequent funding, is the number of papers published by researchers working on various projects. That was the main reason for the Ministry of Science and Technology to initiate a concept of electronic bibliography in the network environment"</u>
Data collection per se	"The main objective is <u>to comprehensively register publication activities of</u> universities in electronic form."

Source: Unpublished results from Linda Sīle's doctoral study. Analysis based on textual materials collected in 2016-2018.

#### Overview of database logics for 12 databases

	A : Enlightenment	B : New Public Management	C : Bibliometric research	D : Data collection per se	E : Other purpose
1 : BFI	0%	100%	0%	0%	0%
2 : COBISS	0%	61,54%	0%	38,46%	0%
3 : CREPČ	20,9%	58,43%	3,8%	16,39%	0,48%
4 : CRISTIN	12,14%	77,08%	0%	10,78%	0%
5 : CROSBI	34,57%	51,08%	2,7%	11,65%	0%
6 : MTMT	41,24%	43,74%	0%	9,85%	5,18%
7 : PBN	13,04%	64,73%	0%	13,04%	9,18%
8 : RINC	47,07%	28,8%	2,39%	12,5%	9,24%
9 : RIV	2,29%	97,71%	0%	0%	0%
10 : SWEPUB	7,48%	54,01%	32,12%	6,02%	0,36%
11 : VABB-SHW	0%	97,24%	0%	2,76%	0%
12 : VIRTA	5,59%	79,89%	2,61%	11,92%	0%

Source: Unpublished results from Linda Sīle's doctoral study. Analysis based on textual materials collected in 2016-2018.

## Example #1

The main purpose of the database is to calculate bibliometric indicators for research funding allocation system

--The database does not have a user interface online for searching, browsing and other ways of interacting with the database content
--The annual reporting of indicators is hindered since there is no information on affiliations of authors for publications recorded in the database

What could be the next step in the development of this database?

## Example #2

The idea is to implement a database for a wide variety of purposes: for research evaluation and funding allocation, for reporting to funders and the government, for knowledge dissemination, for bibliometric research, and for linguistic and historical research.

- --The different potential user groups are involved in discussion on the database design.
- --Each user groups has different requirements for the database.
- --The available resources are limited and uncertain.

What could be the next step in the development of the database?

#### I. Identify and make explicit the purpose(s) of the database

Keeps everyone on the same page



#### 7. Collaborate with stakeholders

Contributes to usability, publicity and quality of the database

Researchers	Users	Governmental organisations	Research performing organisations
Librarians	International organisations	Bibliometricians	Developers
Data providers	Funders	?	?

#### 10. Aim for inclusion of a wide range of research output types

Facilitates multiple uses of the database

Institutional or departmental research evaluation Reports to funders and governmental bodies Research output overviews online Information retrieval

•••

# 10. Aim for inclusion of a wide range of research output types What does a 'wide range' mean?



# 10. Aim for inclusion of a wide range of research output types What does a 'wide range' mean?

#### What is included in the database?

- -- Is the focus on publications?
- -- Can alternative forms of research output be included?

(corpora, artworks, software, research protocol)

- -- Can forms of knowledge dissemination be included?
  - (lectures, conference talks, interviews with the press, organisation of conferences)

# How fine-grained is the classification of research output types?

- -- Will there be one or multiple categories for different kinds of books?
  - (e.g., monographs, edited volumes, and text books)
- -- Will there be separate categories depending on the audience of the publication (scholarly, professional, general public) or the use of peer-review or location of the publisher?



#### 10. Aim for inclusion of a wide range of research output types

The wider the range, the more resources are required

#### At the implementation stage:

-- development of a vocabulary (+ definitions + guidelines for implementation and use)

#### At the use stage:

- -- Responses to questions about the use of different research output
  - -- discussions
- -- Monitoring of the accuracy of the research output types

# 30. Follow and adapt to developments in research practices, research policy, and database maintenance

Ensures that the database remains up-to-date

Changes in the needs of users --different requirements for reporting? --demands for expanded functionality? Changes in research practices --new research output types? --new academic disciplines? Changes in research policy --demands for evidence on different questions Changes in technologies

! Backward compatibility

! Up-to-date documentation

! Procedures for data provenance

## Discussion in groups

I. Consider the kinds databases for which this recommendation is highly important and for which it can be less relevant!

--Is it applicable to all databases for research output? Only for those used in research evaluation? Only for those focused on the social sciences and humanities?

--Does it matter whether the database is in-house built solution or developed using an open source platform?

--How about institutional databases? Institutional repositories? Digital archives? Is the recommendation applicable?

2. Relate the recommendations to the database you have experience with!

-- Is the practice the same as described in the recommendation?

--If no, what would be steps towards implementing the recommendation? What challenges could be anticipated?

- -- If yes, how the aspect in consideration could be developed further?
- --How the implementation of this recommendation could influence user experience?

#### 20 minutes

3. Share a 'take home' message from the discussion in your group!

Thank you!

Linda.Sile@uantwerpen.be



&pid=1

Pictu

#### References

DORA. (2012). San Francisco Declaration on Research Assessment. Retrieved from https://sfdora.org/read/

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature*, 520, 429–431.

Puuska, H.-M., Guns, R., Pölönen, J., Sivertsen, G., Mañana-Rodríguez, J., & Engels, T. (2018). Proof of concept of a European database for social sciences and humanities publications: Description of the VIRTA-ENRESSH pilot (p. 23). Retrieved from CSC & ENRESSH website: <u>https://doi.org/10.6084/M9.FIGSHARE.5993506</u>

Sīle, L., Guns, R., & Engels, T. C. E. (2018, November 13). Towards more consistent, transparent, and multipurpose national bibliographic databases for research output. Retrieved 14 February 2019, from LSE Impact Blog website: <u>https://blogs.lse.ac.uk/impactofsocialsciences/2018/11/13/towards-more-consistent-</u> <u>transparent-and-multi-purpose-national-bibliographic-databases-for-research-output/</u>

Sīle, L., Guns, R., Sivertsen, G., & Engels, T. C. E. (2017). European Databases and Repositories for Social Sciences and Humanities Research Output (p. 25). Retrieved from ECOOM & ENRESSH website: https://doi.org/10.6084/m9.figshare.5172322.v2

Sīle, L., Pölönen, J., Sivertsen, G., Guns, R., Engels, T. C. E., Arefiev, P., ... Teitelbaum, R. (2018). Comprehensiveness of national bibliographic databases for social sciences and humanities: Findings from a European survey. *Research Evaluation*, 27(4), 310–322. <u>https://doi.org/10.1093/reseval/rvy016</u>