overview of the ENRESSH good practices manual for national bibliographic databases

Linda Sîle, doctoral researcher
Centre for R&D Monitoring (ECOOM), University of Antwerp, Belgium

ENRESSH Training school : Poznan, 22/10/2019
Result of collaborative work between:

Linda Sile
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
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)

“11. 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


ENRESSH-VIRTA Proof of Concept

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

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 research publications is checked?
- How collaboration with governmental bodies and/or authors of research output is organised?

The programme of the workshop contains 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@antwerp.be

Organised by ECOOM & University of Antwerp

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
Content
Overview

1. Design
2. Data collection
3. Organisation
4. Research output types
5. Vocabularies, authority control and identifiers
6. Quality control
7. Data use
8. Transparency and sustainability
1. 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
1. Identify and make explicit the purpose(s) of the database

| Helps to design the database in line with users’ needs |

- Hi mom, what are you drawing?
- A cartoon illustrating that you first need to understand user needs before building tools for them.
- Well duh, isn't that obvious?!
- You would think so...

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. The portal functions of the [database] can lead them to scientific results - publications residing in repositories - they can understand and use in their studies.”

**New Public Management**

“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”

**Data collection per se**

“The main objective is to comprehensively register publication activities of … universities in electronic form.”

### Overview of database logics for 12 databases

<table>
<thead>
<tr>
<th></th>
<th>A: Enlightenment</th>
<th>B: New Public Management</th>
<th>C: Bibliometric research</th>
<th>D: Data collection per se</th>
<th>E: Other purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 : BFI</td>
<td>0%</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>2 : COBISS</td>
<td>0%</td>
<td>61.54%</td>
<td>0%</td>
<td>38.46%</td>
<td>0%</td>
</tr>
<tr>
<td>3 : CREPČ</td>
<td>20.9%</td>
<td>58.43%</td>
<td>3.8%</td>
<td>16.39%</td>
<td>0.48%</td>
</tr>
<tr>
<td>4 : CRISTIN</td>
<td>12.14%</td>
<td>77.08%</td>
<td>0%</td>
<td>10.78%</td>
<td>0%</td>
</tr>
<tr>
<td>5 : CROSBI</td>
<td>34.57%</td>
<td>51.08%</td>
<td>2.7%</td>
<td>11.65%</td>
<td>0%</td>
</tr>
<tr>
<td>6 : MTMT</td>
<td>41.24%</td>
<td>43.74%</td>
<td>0%</td>
<td>9.85%</td>
<td>5.18%</td>
</tr>
<tr>
<td>7 : PBN</td>
<td>13.04%</td>
<td>64.73%</td>
<td>0%</td>
<td>13.04%</td>
<td>9.18%</td>
</tr>
<tr>
<td>8 : RINC</td>
<td>47.07%</td>
<td>28.8%</td>
<td>2.39%</td>
<td>12.5%</td>
<td>9.24%</td>
</tr>
<tr>
<td>9 : RIV</td>
<td>2.29%</td>
<td>97.71%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>10 : SWEPUB</td>
<td>7.48%</td>
<td>54.01%</td>
<td>32.12%</td>
<td>6.02%</td>
<td>0.36%</td>
</tr>
<tr>
<td>11 : VABB-SHW</td>
<td>0%</td>
<td>97.24%</td>
<td>0%</td>
<td>2.76%</td>
<td>0%</td>
</tr>
<tr>
<td>12 : VIRTÅ</td>
<td>5.59%</td>
<td>79.89%</td>
<td>2.61%</td>
<td>11.92%</td>
<td>0%</td>
</tr>
</tbody>
</table>

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?
1. 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

<table>
<thead>
<tr>
<th>Researchers</th>
<th>Users</th>
<th>Governmental organisations</th>
<th>Research performing organisations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Librarians</td>
<td>International organisations</td>
<td>Bibliometricians</td>
<td>Developers</td>
</tr>
<tr>
<td>Data providers</td>
<td>Funders</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>
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?

Source: Multiple sources consulted in May 2018
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

1. 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?

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
References


