

Entanglement of bibliographic database content and data collection practices

Rethinking data integration using findings from a study of European databases for research output

Linda Sile, PhD Candidate in Social Sciences



University
of Antwerp

ECOOM

Expertisecentrum O&O Monitoring

Outline

Background

European Network for Research Evaluation in the Social Sciences and Humanities

Study on databases for research output in the social sciences and humanities

Challenges in making sense of the database content

Implications for data integration



ENRESSH

European Network for Research Evaluation

in the Social Sciences and Humanities

COST Action 15137

Communication in Europe

23 April 2018

30252

20 April 2018

Working with national bibliographic databases for research output

16 April 2018

Becoming Metric-Wise

Goal: (a roadmap for) European bibliographic database for research output in the social sciences and humanities

... created by integrating currently existing databases

Challenge: limited information on currently existing databases

Solution: **Launch of a study of national bibliographic databases in Europe (in August 2016)**

A study on national bibliographic databases for research output in the social sciences and the humanities in Europe

- 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
- Participant observation** *thick description*
of the content of databases and
of the process of databases being surveyed

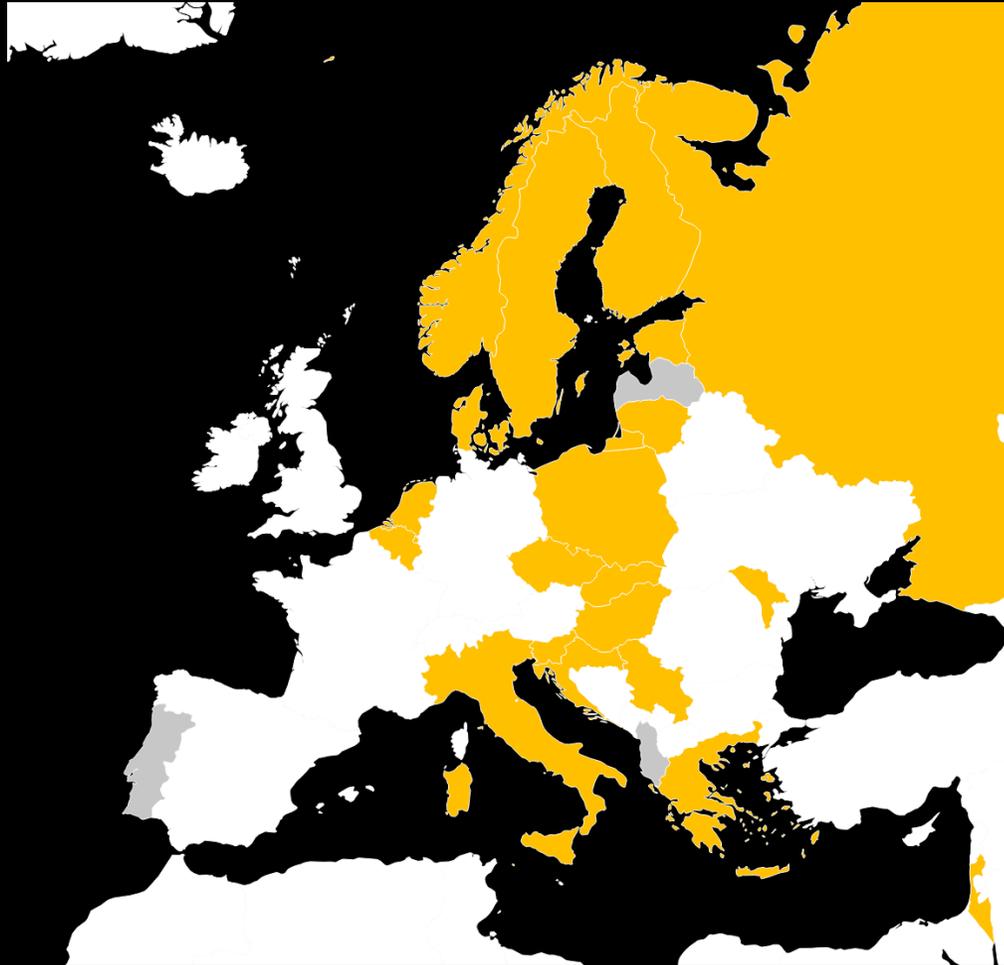
ECOOM
Expertisecentrum O&O Monitoring

cost
EUROPEAN COOPERATION
IN SCIENCE AND TECHNOLOGY



European databases and repositories for Social Sciences and Humanities research output

Report
July 2017



- National database identified and described
- National database is being implemented

The number of databases that include the most common types of publications



Outline

Background

European Network for Research Evaluation in the Social Sciences and Humanities

Study on databases for research output in the social sciences and humanities

- ➔ Challenges in making sense of the database content
- Implications for data integration

category = category

book chapter = book chapter

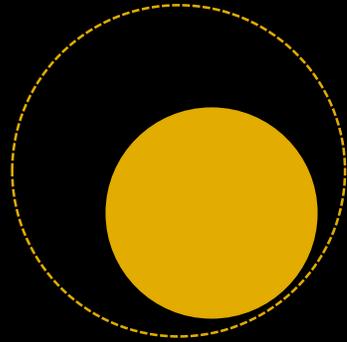
~~category = definition = category~~

~~book chapter = independent part of a monographic publication or an editorial collection (excluding articles in conference proceedings) = book chapter~~

book chapter = book chapter?

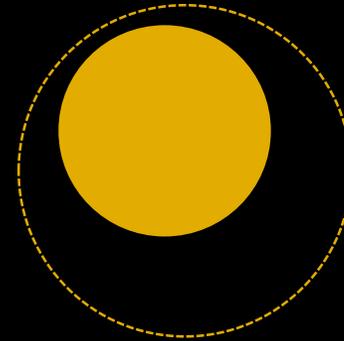
Database A

any book chapter except for
those shorter than 4 pages



Database B

any book chapter except for
introductions, prefaces,
forewords, afterwords, etc.



book chapter = book chapter?

category = category

category = definition = category

category = practice = category

the content of bibliographic databases is entangled with data collection practices

Outline

Background

European Network for Research Evaluation in the Social Sciences and Humanities

Study on databases for research output in the social sciences and humanities

Challenges in making sense of the database content

➔ Implications for data integration

research
practice



Who enters data? How?

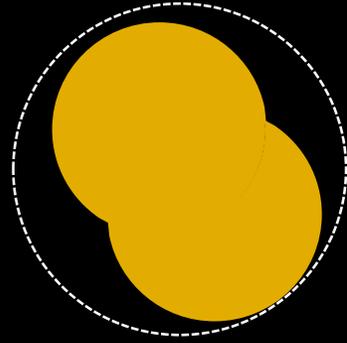
What processing is carried out?

*Are metadata assigned from external sources?
If so, from which sources and how?*

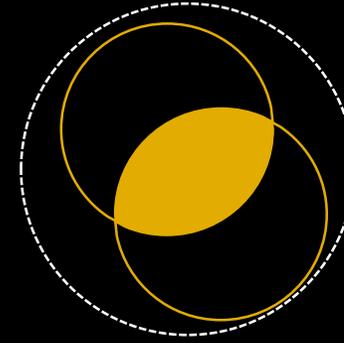
*Are there any records removed?
If so, which ones?*

Research
Information
System

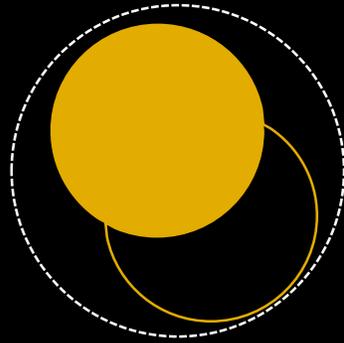
Integrated solution #1
differences assumed irrelevant



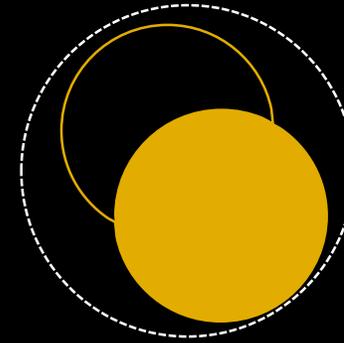
Integrated solution #2
intersection of standards



Integrated solution #3
Database A = standard



Integrated solution #4
Database B = standard



What are the implications for data integration when data are drawn from different national contexts?

Problem #1:

It is time-consuming to be aware of differences in practices

Problem #2:

Choosing a standard that leads to exclusion of records from *some* contexts leads to a bias



Transparency

Towards being explicit about **what is and is not known**

-- in databases *to be* integrated

-- in databases created by integrating data

Reflexivity

Towards awareness of the value-guided aspects of data integration

In more practical terms

Transparency

Detailed documentation of

- databases *to be* integrated
- databases created by integrating data

Reflexivity

Original metadata remain with records after integration

All records included

- quality-related metadata added to records without altering the content of an integrated solution



Thank you!

For more information:

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

Sile, L., Pölönen, J., Sivertsen, G., Guns, R., Engels, T. C. E., Arefiev, P., ... Teitelbaum, R. (in press). Comprehensiveness of national bibliographic databases for social sciences and humanities: findings from a European survey. *Research Evaluation*.

Linda.Sile@uantwerpen.be

Workshop in Antwerp, September 10-11

Registration by June 30

 Working with national bibliographic databases for research output
Antwerp, September 10-11

Participation Information

You are here: [UAntwerpen](#) > [Centra](#) > Expertisecentrum Onderzoek en Ontwikkelings...

Working with national bibliographic databases for research output

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

Further information: www.uantwerpen.be/ecoom