Entanglement of bibliographic database content and data collection practices

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

Linda Sīle, PhD Candidate in Social Sciences

CRIS 2018: Umeå, June 14-16, 2018
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
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
thick description
observation
of the content of databases and
of the process of databases being surveyed
National database identified and described

National database is being implemented

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

- Journal articles
- Monographs
- Anthologies/edited books
- Book chapters
- Conference proceedings
- Doctoral theses

Number of databases

0 5 10 15 20 25

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


Linda.Sile@uantwerpen.be

Workshop in Antwerp, September 10-11
Registration by June 30

Further information: www.uantwerpen.be/ecoom