DZHW

German Centre for Higher Education Research and Science Studies

The Research Core Dataset: standardizing the collection and provision of research information

Training school "National bibliographic databases and their uses for evaluating and understanding research"
Poznań – 22 October 2019

Sophie Biesenbender

1.1 Background

- Research information: numerical information on research activities and outputs (of individual researchers, departments, institutions etc.)
  - research information ≠ research data
- **Research information (RI)** used for institutional reporting (to ministries, funding organisations, official statistics etc.), planning and evaluation processes (e.g. internal controlling), outreach and communication
- **RI standards** needed
  - for research institutions to reduce efforts with the collection and processing of RI
  - to ensure data quality and comparability across research institutions and to enable evidence-based policy making

1.2 The German science system

- Collection and processing of research information particularly fragmented in the German science system
  - distributed legislative and regulatory authority for the governance of research institutions (Federal and State Governments)

- New public management (NPM) since the 1980s
  - growing autonomy of institutions
  - growth of third-party and performance-based funding
  - increasing relevance of rankings and need for transparency

1.2 The German science system (ctd.)

- Increasing need for reporting of research information (in Germany)
  - external: rankings, ratings, (output-oriented) reporting to ministries and funding organizations, Statistical Offices, Council of Science and Humanities, implementation of performance-based funding systems
  - internal: development of internal governance processes, evaluation, controlling, management, communication

2.1 RCD development process

- Recommendations of the German Council of Science and Humanities (2013)
- Specification project for standards development
  - Project period: 2013-2015
  - Coordinator: iFQ (predecessor of DZHW Berlin)
  - Structure: four expert groups, partners: pilot institutions, non-university research institutions, scientific societies
  - Development process:
    - Development of first draft of RCD specification by expert groups in coordination with partners (*alpha release*)

2.1 RCD development process (ctd.)

- Evaluation and modification of draft specification by advisory board (working group of the Science Council) (*beta release*)
- Publication of beta release and open invitation to comment
- Revision of specification according to the comments (*first release 1.0*)

- Recommendations of the German Council of Science and Humanities (2016)
  - implementation of RCD specification (by research institutions and report-requesting organisations)
  - establishment of a helpdesk to facilitate implementation process

2.2 RCD structure and contents

- Principles of the RCD
  - not a (central) dataset in the strict sense (but a standard specification for the decentralized reporting and processing of RI)
  - does not create new reporting obligations for research institutions but intended to be used for existing ones
  - development of a “flexible” specification that can be customized for different reporting purposes
  - different layers: core-shell structure (1. layer: mandatory information, 2. layer: optional information, 3. layer: areas for further development)

2.2 RCD structure and contents (ctd.)

- RCD specification contains standard definitions for RI in six areas
  - Employed staff
  - Young researchers (including structured doctoral programs)
  - Third-party funding and budget
  - Patents and spin-offs
  - Publications
  - Research infrastructures
- Aggregate vs. basic data
- Different parameters and aggregation levels (flexible use)

2.2 RCD structure and contents (ctd.)

- Employed staff & Young researchers
  - **Full-time equivalents & Number of staff**
    - Parameters: gender, staff category, form of financing, highest academic degree, type of contract (permanent vs. fixed-term), citizenship, age group, type of job (full-time, part-time)
    - Aggregation levels: organisational unit, discipline, research field
  - **Number of doctoral students**
    - Parameters: gender, citizenship, start of dissertation (year), supervisor, structured doctoral programme, cooperation with other research institutions, country of graduation (PhD qualification)
    - Aggregation levels: organisational unit, discipline, research field

2.2 RCD structure and contents (ctd.)

- Publications
  - **Lists of publications**
    - Parameters: publication type, document type, author, title, publisher, source, identifier, format, language, access rights, peer-reviewed, thesis, resource, funding, grant number
    - Aggregation levels: organisational unit, discipline, research field

2.2 RCD structure and contents (ctd.)

- Publications
  - **Lists of publications**
    - Parameters: publication type, document type, author, title, publisher, source, identifier, format, language, access rights, peer-reviewed, thesis, resource, funding organization, grant number
    - Aggregation levels: organisational unit, discipline, research field

2.2 RCD structure and contents (ctd.)

- Definition of “publication”
  - A scientific publication is a recorded and published product resulting from the research of one or more persons. A publication must correspond to an independent intellectual achievement; purely formal compilations should not be regarded as scientific publications. Scientific publications are directed mostly to research colleagues, but formats in which scientific results are communicated to a wider public, should also be considered. Publications should be permanently available. Citable and published research data are also considered as publications.

2.2 RCD structure and contents (ctd.)

- Development of publication types
  - Overview of existing publication formats in German repositories
  - Selection of relevant formats by the expert group
  - Consultation of 259 scientific associations (covering all disciplines) and 4 four pilot institutions
  - Subsequent revision of the classification of publication types
  - Selection process includes discussion of metadata standards WoS, RDA, CRIStin and CERIF
  - Result: modified Dublin Core

2.2 RCD structure and contents (ctd.)

- Publication types
  - Book
    - Monograph
    - Edited book
    - Bibliography
    - Scholarly edition of primary sources
  - Article
    - Journal article
    - ePaper
    - Chapter in edited book
  - Software
  - Special issue
  - Research data
  - Conference poster
  - Contribution/interview related to scientific activities published in mass media
  - Slides of presentations
  - Working paper
  - *New publication type*

2.2 RCD structure and contents (ctd.)

- Document types to further differentiate “Articles” (publication type)
  - Editorial
  - Scientific article
  - Overview article (literature overview article)
  - Bibliography
  - Review
  - Scholarly edition of primary sources
  - Letter to the editor
  - Meeting abstract

3 Discussion and questions

- RCD principles:
  - Decentral systems/repositories vs. central database
  - Harmonization of institutional systems/repositories vs. creation of a national/central research information system
    - → Data ownership remains with research institutions
    - → Compatible with heterogeneous legal systems (federated system)
  - Challenge: Quality control, establishment of harmonized procedures (comparability of the data)

3 Discussion and questions

- Challenge:
  - Find definitions and classifications that are useful (i.e. that reflect research (activities and outputs)) AND
  - standardize (many) existing information needs (over different types of institutions and disciplinary foci)
    - How to anticipate/identify information needs?
    - How to find a trade-off between conflicting needs/purposes?
    - How to reflect disciplinary characteristics?
    - Define priority of existing metadata standards (RDA, Dublin Core, CERIF…)

3 Discussion and questions

- A carefully designed expert-guided development process, but do we need more?
  - involving as many stakeholders as possible
  - gradually increasing transparency of the development process (from closed expert groups to consultations of academic societies to public request for comments) *[process legitimacy]*
  - open documentation and visualization of the development process, the discussions and the pros and cons of single decisions and definitions *[output legitimacy]*

3 Discussion and questions

- How to evaluate the quality of the standard (e.g. the typology of publications and other classifications)?
  - Systematic assessment of information requirements and reporting purposes
  - Assess whether the standard allows for the construction of useful indicators
- How to assess the quality of implementation and hence the quality/comparability of data?
  - Assessment of institutional technical and organizational solutions
  - Establish and diffuse good practices and recommendations regarding processes of quality control and validation as well as technical solutions

Thank you very much for your attention!

Further information (English version to be online soon):
http://www.kerndatensatz-forschung.de/

Dr. Sophie Biesenbender
German Centre for Higher Education Research and Science Studies
Schützenstraße 6a | 10117 Berlin | www.dzhw.eu | Germany
eMail: biesenbender@dzhw.eu