KonsortSWD: A New Joint German Research Data Infrastructure for the Social Sciences

Christof Wolf
GESIS – Leibniz-Institute for the Social Sciences

ESRA 2021, virtual, 23.07.2021
Contents

- NFDI – some context
- Strong Starting Position – current Social-Science RDM-infrastructure
- Still Room to Grow – objectives
Context: Research Data Management

Research Data Management

- is key to *re-using data* and *replicating* research results
- is key to generating *new knowledge* (based on existing data)
- is often considered to be either fiendishly *difficult* or *boring*
Key challenges for RDM are

- heterogeneous research cultures
- incompatible existing data silos
- lack of standards
- lack of skills
- lack of incentives
- combined with emergence of new data types
Context: NFDI

The German National Research Data Infrastructure

- is more than a national EOSC
- because it does not start technically (i.e. tools, software)
- but rather builds on RDM needs of research communities
Context: NFDI

- The aim of the NFDI is to systematically manage scientific and research data, provide long-term data storage, backup and accessibility, and network the data both nationally and internationally.

Context: NFDI

- NFDI responds to similar international activities, e.g.
  - ODISSEI (Open Data Infrastructure for Social Science and Economic Innovations) - Netherlands
  - National Collaborative Research Infrastructure Strategy – Australia
  - Infrastructure of national interest – Sweden

- and will interface with EOSC and RDA
Context: NFDI

- Up to 30 consortia from all disciplines representing research communities
- Selected in 3 rounds (2020-2022)
- €85m / annually
- base for data-type-specific and general services to facilitate RDM and make data re-usable across domains.
Strong Starting Position in the Social Sciences

- Existing infrastructure of almost 40 research data centers
- covering mostly surveys and administrative data
- have established quality assurance
- strong research based output
Research Data Centers (RDCs)
Key figures of RDCs in 2019

- **4,371** Datasets
- **2,359** Publications based on RDC data
- **431 datasets added**
- **55,270** External Data users
- **294** Full-time equivalent staff
- **79,878** Downloads of open access datasets
- **521** Scientific publications of RDC staff

**Publications** based on RDC data
Room to Grow: KonsortSWD Objectives

- **Strengthen** the FAIRness of (meta)data
- **Widen** the scope of data available through Research Data Centres (RDC)
- **Deepen** the cooperation between RDCs
KonsortSWD

- Improve **findability** on mainstream search engines
- Focus on **interoperability** – not only on a technical level (e.g. harmonization of core concepts)
Encourage creation of more RDCs for data as yet unavailable for research

Focus on new data types (particularly text and qualitative)

Investigate potential for data linkage
KonsortSWD

- Create PIDs at sub-study level
- Harmonize processes for data-access
- Provide central services (like remote access) to assist RDC
# Objectives: Our Services and Two Examples

| Service                                      | Description                                                                 |
|----------------------------------------------|                                                                            |
| Automatic coding of open questions           |                                                                            |
| Actively involve professional associations   |                                                                            |
| Q&A Forum for RDC users                     |                                                                            |
| RatSWD (German Data Forum)                  |                                                                            |
| Open Metadata & Data format                 |                                                                            |
| Expand network of partners                  |                                                                            |
| Improve findability (Google et al.)         |                                                                            |
| Quality assurance for RDC                   |                                                                            |
| RDM Grants                                   |                                                                            |
| Linking Text and other data types           |                                                                            |
| PID Service below study level               |                                                                            |
| Unified points of access to RDC             |                                                                            |
| Training for RDM Personell                  |                                                                            |
| federated archiving qualitative data         |                                                                            |
| Harmonization of core variables             |                                                                            |
| Support for RDC (CoreTrust Seal, …)         |                                                                            |
| Interfaces for Data exchange                |                                                                            |
| Support for qualitative RDM                 |                                                                            |
- Survey data are rarely interoperable as surveys measure concepts with different instruments
- We support questionnaire design with best practice guidelines
- For latent constructs (e.g. attitudes, interest, or values), we create a harmonization database
- Will go online with “political interest” during 2021
Open and semi-open response formats are part of most surveys

We develop a tool for efficient and quality-assured coding (focus: occupations, education)

Allows coding and validation with help of automatically generated suggestions, commenting and process documentation

A beta version of CODI will be available in July 2022
Thank You

Great collaboration starts small

KonsortSWD is funded by Deutsche Forschungsgemeinschaft (DFG) as part of NFDI - 442494171