

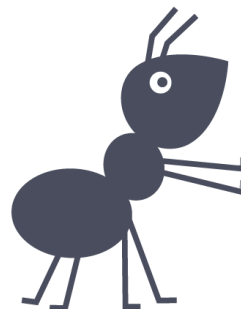
KonsortSWD



Consortium for the
Social, Behavioural, Educational
and Economic Science

ESRA 2021,
virtual, 23.07.2021

nfdi



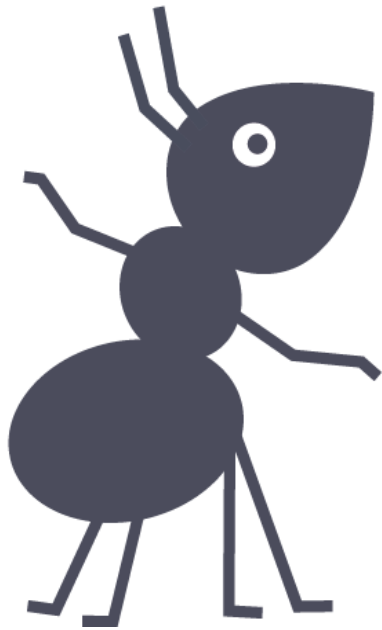
KonsortSWD: *A New Joint German Research Data Infrastructure for the Social Sciences*

Christof Wolf

GESIS – Leibniz-Institute for the Social Sciences

Contents

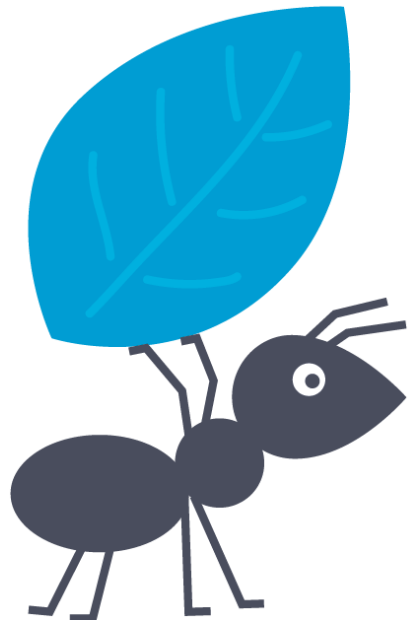
- NFDI – *some context*
- Strong Starting Position – *current Social-Science RDM-infrastructure*
- Still Room to Grow – *objectives*



Context: Research Data Management

Research Data Management

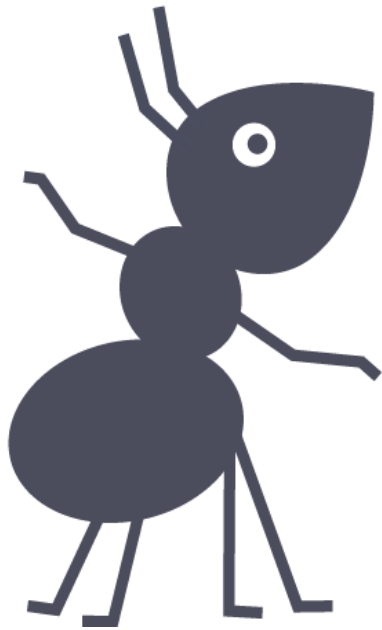
- is key to **re-using data** and **replicating** research results
- is key to generating **new knowledge** (based on existing data)
- is often considered to be either fiendishly **difficult** or **boring**



Context: Challenges for RDM

Key challenges for RDM are

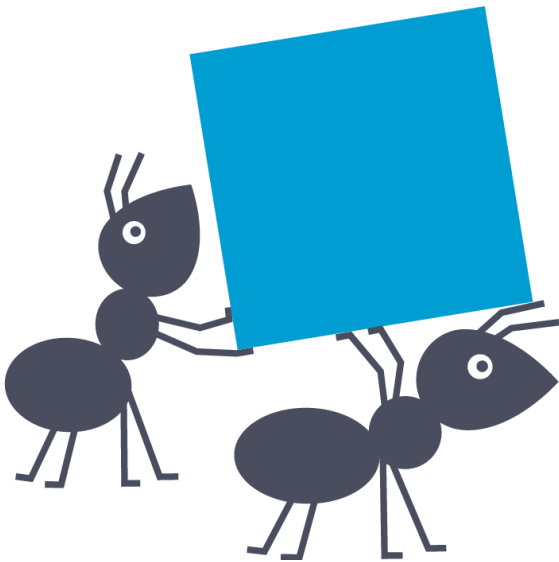
- heterogeneous **research cultures**
- incompatible existing **data silos**
- **lack** of **standards**
- lack of **skills**
- lack of **incentives**
- combined with emergence of **new data types**



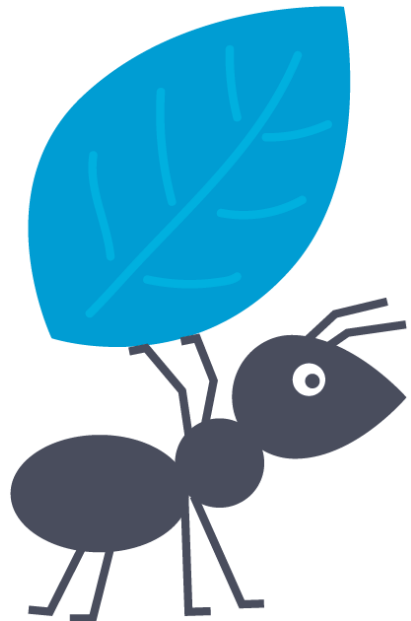
Context: NFDI

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



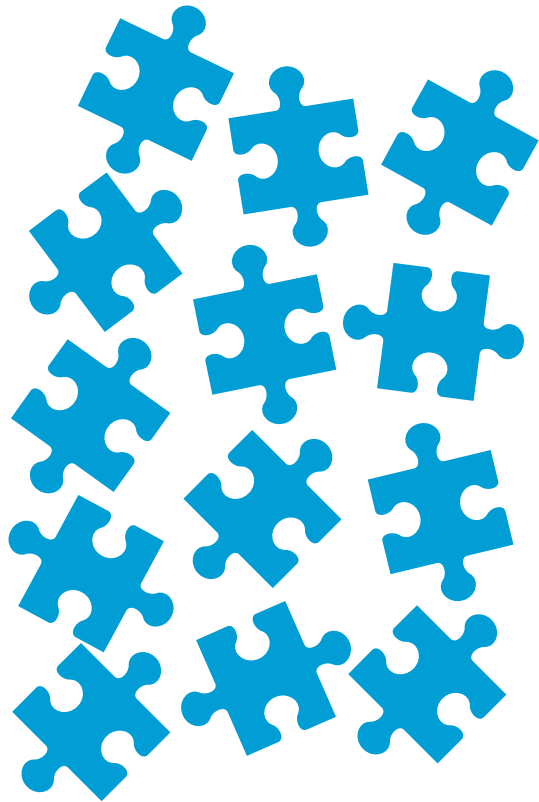
- The aim of the NFDI is to systematically manage scientific and research data, provide long-term data storage, backup and accessibility, and network the data both nationally and internationally.

- https://www.dfg.de/en/research_funding/programmes/nfdi/index.html

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

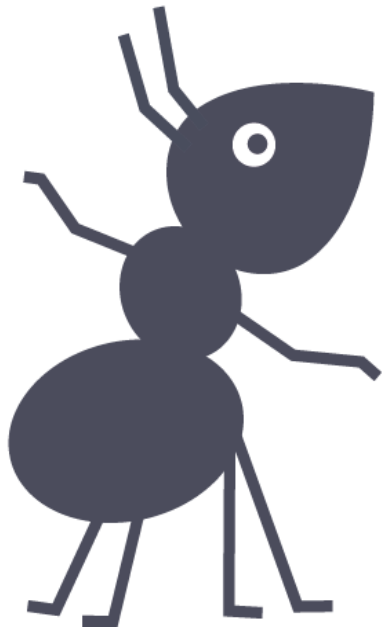
Context: NFDI




- 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)

 bibb Federal Institute for Vocational Education and Training Research Data Centre	 @gesis fdz ALBUS	 DEUTSCHE BUNDESBANK EUROSYSTEM	 AGD	 IWH ForschungsDatenZentrum	 @gesis fdz PIAAC	 SHARE Survey of Health, Ageing and Retirement in Europe 50+ in Europe	 LEIBNIZ INSTITUTE S A F E Sustainable Architecture for Finance in Europe
 eLabour	 IAB RESEARCH DATA CENTRE (FDZ) of the German Federal Employment Agency (IAB) at the Institute for Employment Research (IAB)	 BZgA	 GePaRD German Pharmacological Research Database	 IDSC INTERNATIONAL DATA SERVICE CENTER	 leibniz-psycho.org	 SOEP The Socio-Economic Panel	 @gesis fdz International Survey Programs
 DeZIM .fdz FORSCHUNGS DATEN ZENTRUM	 Kraftfahrt-Bundesamt	 DJI Deutsches Jugendinstitut	 @gesis fdz German Microdata Lab	 STATISTISCHE ÄMTER DES BUNDES UND DER LÄNDER FORSCHUNGSDATENZENTREN	 Qualiservice data sharing	 @gesis fdz Elections	 ZEW FDZ ResearchDataCentre
 LMU-ifo Economics & Business Data Center	 Forschungsdatenzentrum Betriebs- und Organisations DATEN	 DZA German Centre of Gerontology	 forschungsdatenzentrum bildung	 LfBi LEIBNIZ INSTITUTE FOR EDUCATIONAL TRAJECTORIES	 ROBERT KOCH INSTITUT	 FORSCHUNGS-DATENZENTRUM	
 Deutsche Rentenversicherung Bund	 DIW STATIS FDZ	 fdz.DZHW. German Research Data Centre for Higher Education Research and Science Studies	 IQI: Institut zur Qualitätsentwicklung im Bildungswesen Forschungsdatenzentrum	 pairfam Panel Analysis of Intergenerational Relationships and Family Dynamics	 fdz ruhr am rwi	 Leibniz Institute of Ecological Urban and Regional Development	

Key figures of RDCs in 2019



4,371

Datasets

431 datasets added



55,270

External
Data users



79,878

Downloads

of open access datasets



2,359

Publications

based on RDC data



294

Full-time
equivalent staff

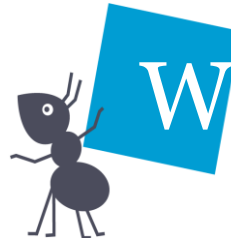


521

Scientific
publications
of RDC staff

Room to Grow: KonsortSWD Objectives

 **S**trengthen
the FAIRness
of (meta)data

 **W**iden
the scope of data
available through
Research Data Centres
(RDC)

 **D**eepen
the cooperation
between RDCs

KonsortSWD



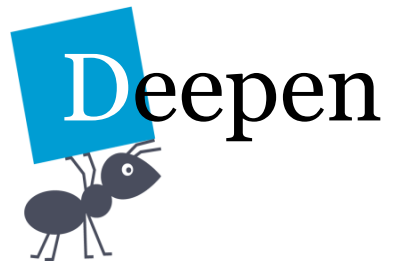
- Improve **findability** on mainstream search engines
- Focus on **interoperability** – not only on a technical level (e.g. harmonization of core concepts)

KonsortSWD



- 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

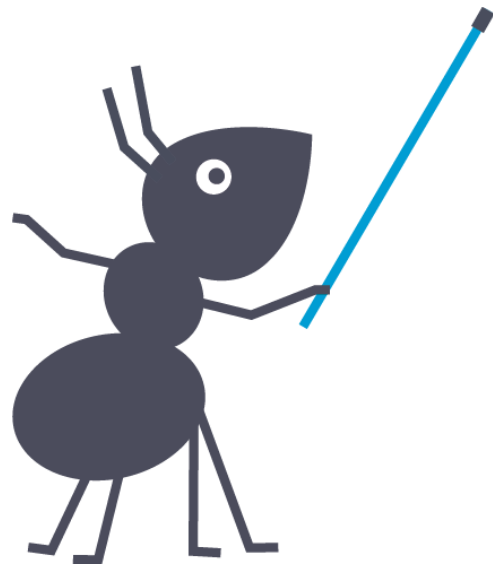
KonsortSWD



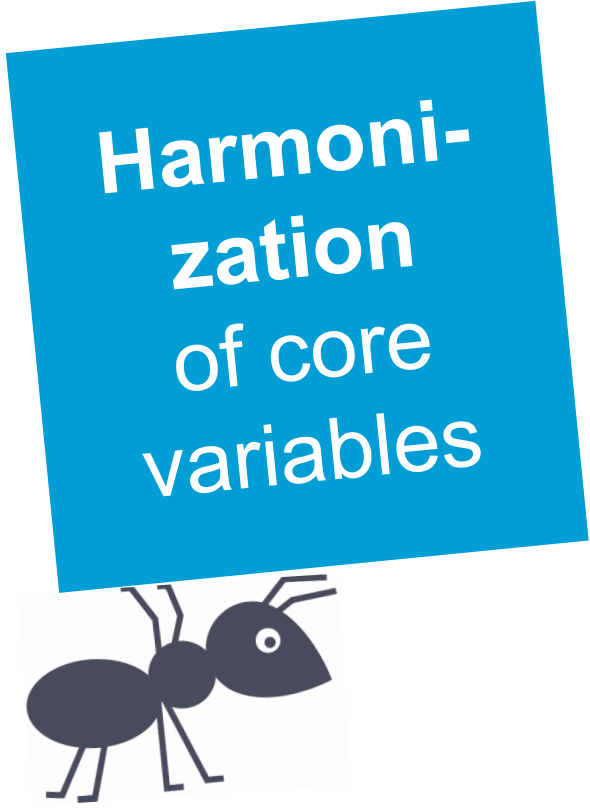
?

- 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



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



Harmonization of core variables

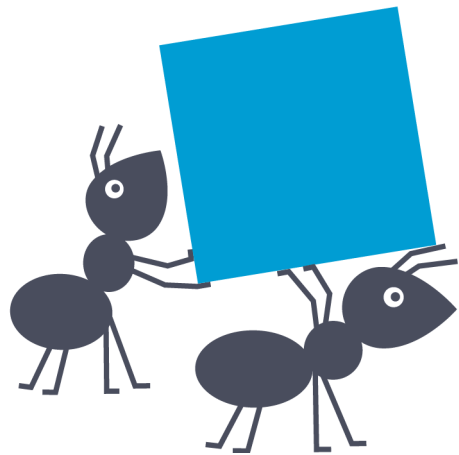
- Survey data are rarely **interoperable** as surveys measure concepts with different instruments
- we support questionnaire design with **best practice guidelines**
- For latent constructs (e.g. attitudes, interest, or values), we create a **harmonization database**
- will go online with “political interest” during 2021



Automatic
coding of
open
questions

- 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](#)

