

Research Data Centers: Enabling Access to Sensitive Social and Economical Data

Thomas Runge

Research Data Alliance:
9th Plenary Meeting, April 05th, 2017



German Data Forum: Key Facts



- Advisory council to the federal government
- 16 members: 8 data producers / 8 data users from research
- Development of a research data infrastructure for the social, behavioral, and economic sciences
- Accreditation of 30 research data centres
- Facilitating access to high-quality data
- Result of independent initiatives from within the scientific community

Social and economic microdata: challenges and characteristics

- 1) Microdata containing information about natural or legal persons: various legal bases and data protection regulations have to be taken into account
➔ **Respect for rights of the data subject**
- 2) This also includes data not generated for research purposes with specific legal requirements (e.g. social law, usability outside Germany)
➔ **Facilitating legal data use for researchers**

Social and economic microdata: challenges and characteristics

- 3) By removing information for the purposes of anonymisation, information content and thus value for research analysis is reduced (particularly for qualitative data)

➔ **Enable maximum analysis potential**

- 4) Increasing methodological heterogeneity (collection methods, data types) and complexity (longitudinal, multi-level, mixed-methods)

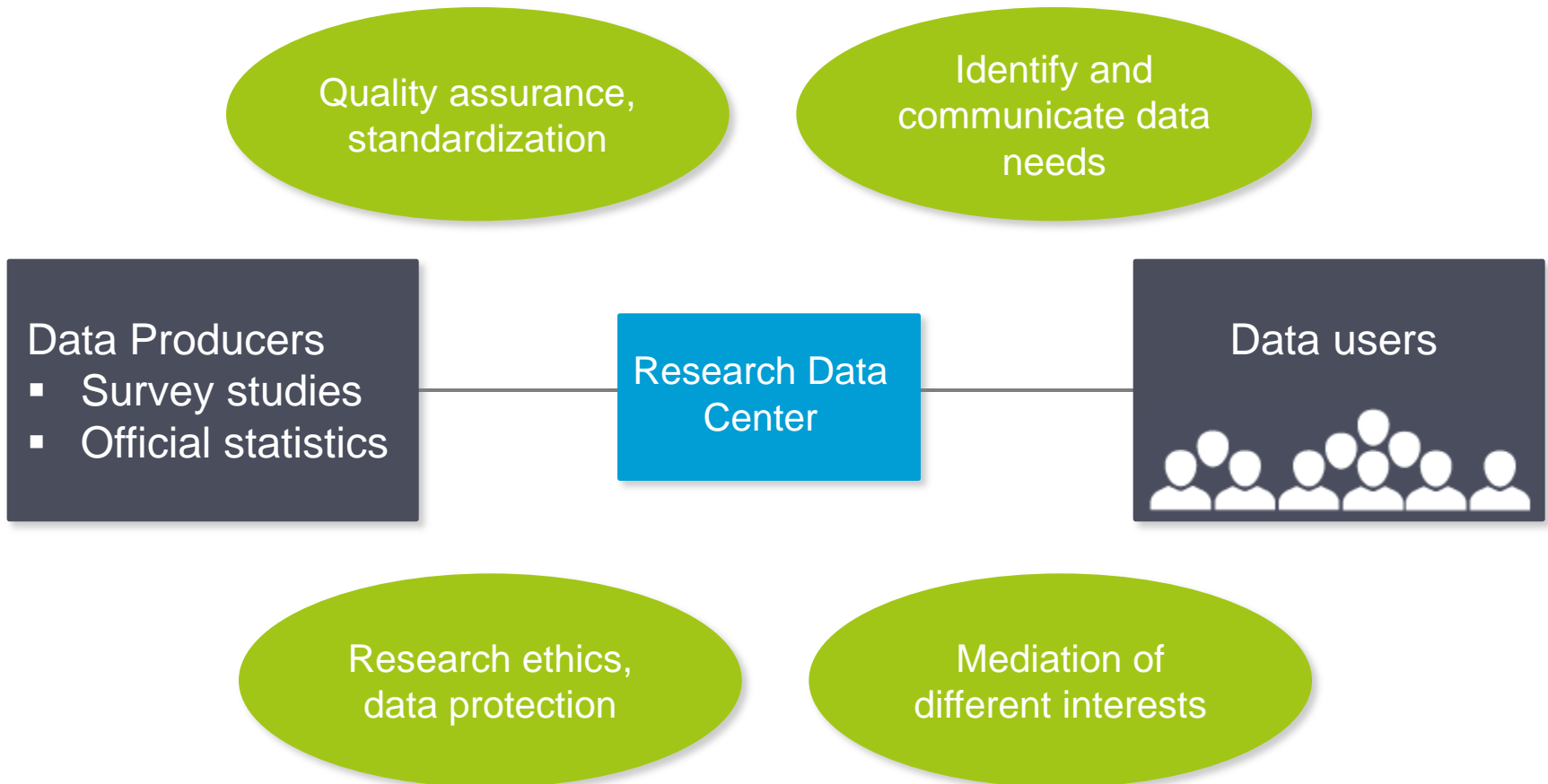
➔ **Quality assurance at all stages**

Data Access Solution: Research Data Centers (RDCs)



- Decentralised data access
- User Support
(documentation, training, guidance)
- Data processing and curation

The Accreditation of RDCs



A Network of Research Data Centers

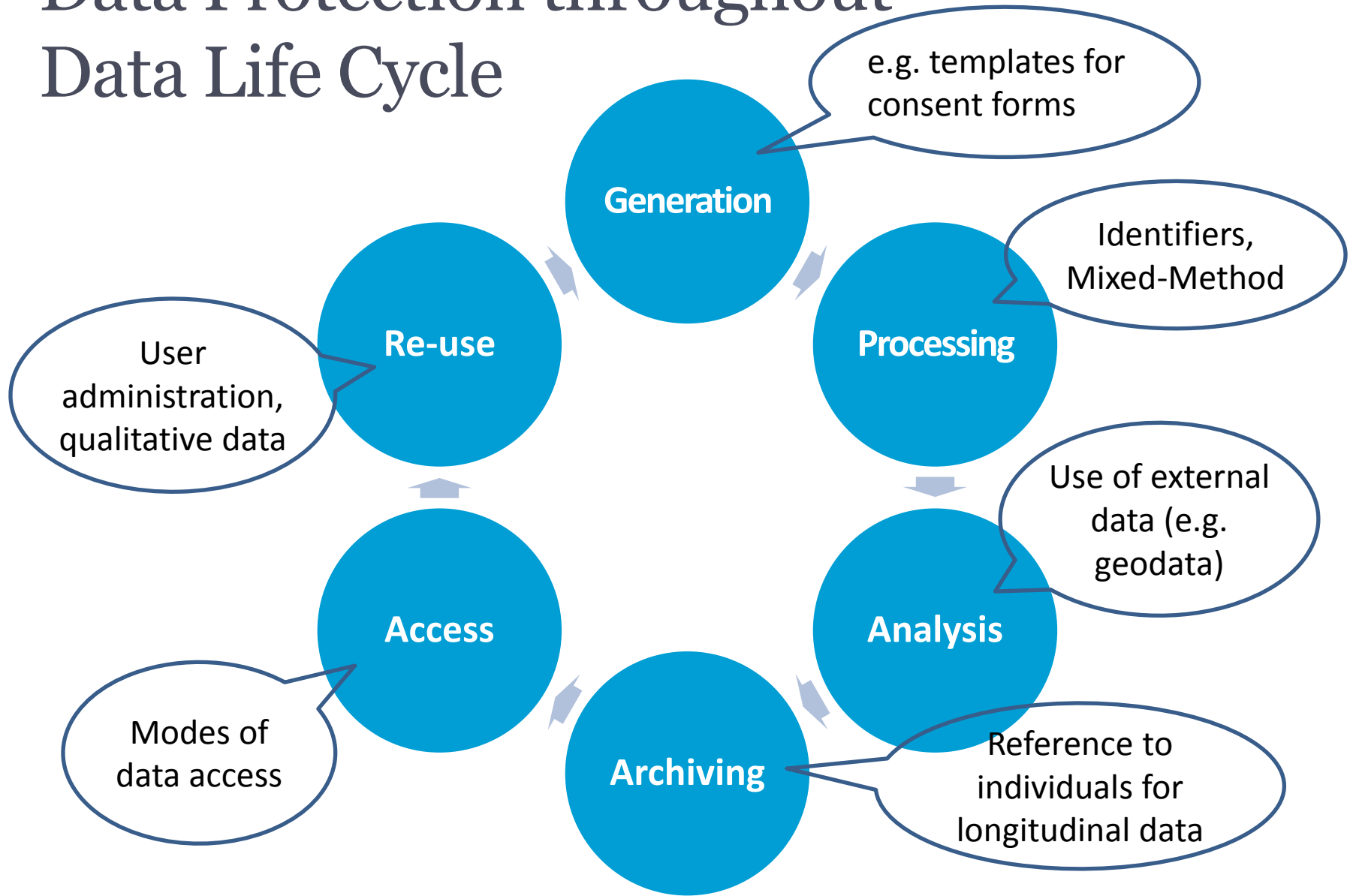
German Data Forum created the **Standing Committee Research Data Infrastructure (FDI)** in 2009, one seat for every research data center accredited by the RatSWD:

- Facilitating synergies
- Exchange of experiences and cooperation
- Common standards on operational level
- Concepts for data protection and data security

30 RDCs Accredited by the RatSWD



Data Protection throughout Data Life Cycle



Technical and Organizational Measures

On the user side:

- Contractual agreements (e.g. prohibiting re-identification)
- Authentication: Limiting data use to research projects

On the RDC side:

- **Pseudonymisation:** separate storage of names/adresses and microdata
- **Anonymisation**
 - Removal or coarsening of variables with high reidentification potential, e.g. geodata, age
 - Drawing of sub-samples or aggregation of variables
 - the more restrictive the data access mode, the weaker the anonymisation concepts

Basic Concepts of Data Access

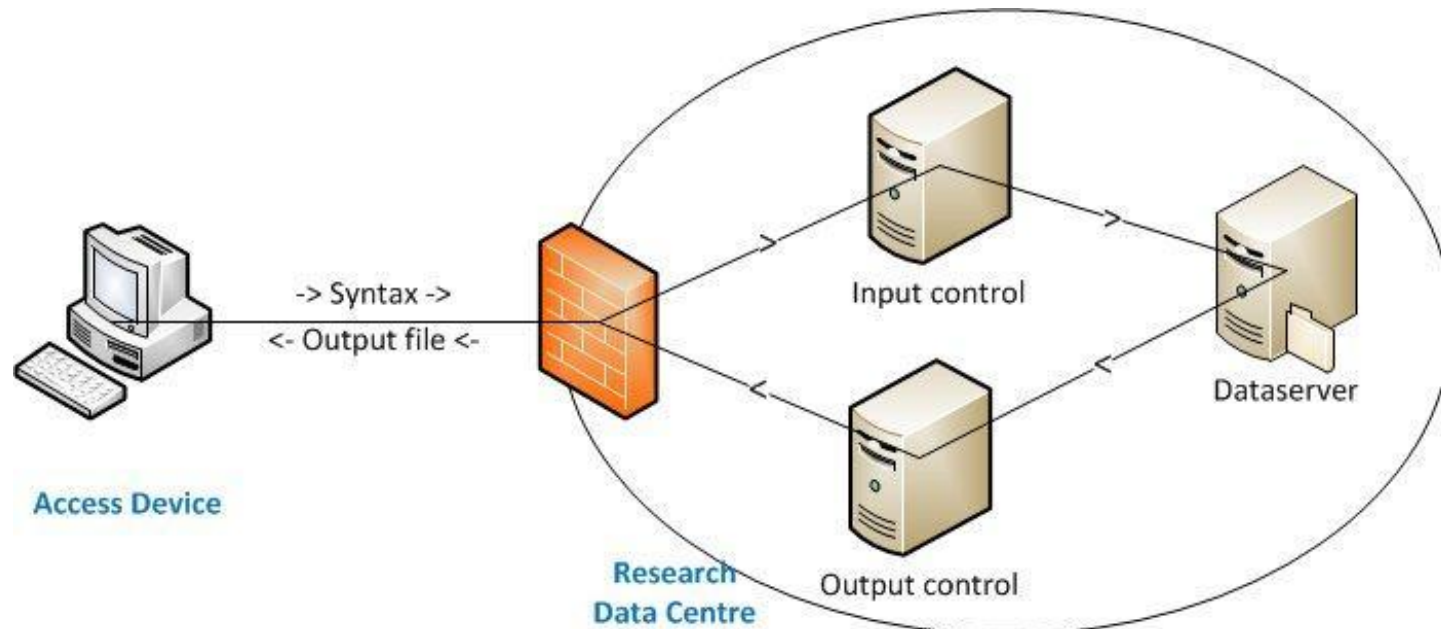
- **Public use file** (Open Access): Reduction of the re-identification risk by providing a significantly reduced amount of data / or by means of modifying data
- **Scientific use file**: Reduced data volume and special security measures on the user side (e.g. contractual agreements & technical precautions)
- **On-site use**: Access to data (usually pseudonymised) only from a special guest workplaces located at the RDC; output is being checked before its release

New Forms of Data Access

- **RDC-in-RDC**: data access ("on-site" use) via other RDC (e.g. data of the Federal Agency for Labor accessible via local branches of the RDC of the statistical offices)
- Establishment of **external sites** in the US and UK (RDC of the Federal Agency for Labor, IAB)
- Authentication of users via personal login in combination with certified **biometric authentication** ("keystroke biometrics"), for example, at the RDC of the National Educational Panel Study (RemoteNEPS)

Remote Execution

Data user sends analysis script to RDC, which is run locally against highly sensitive microdata. After being checked by the RDC, output data is sent back to user. (e.g. LISSY at SOEP; JoSuA at IZA)



Looking Ahead

- Strengthening of federated networks
- International cooperation and cross-border data access
 - ➔ Facilitated by the harmonised European Data Protection legislation
- Paradigm shift in funding policy (long-term, sustainable, focus on networks and platforms)
- International Research Data Centers (ERICs)

Thank you!

Credits:

- Business Office of the German Data Forum
- RatSWD (2017): Activity report of the accredited Research Data Centers of the German Data Forum for the year 2015
- Stefan Bender (2016): Presentation at the Public-Academic Research Colloquium
- Stefan Liebig (2016): Presentation at the RDA-DE Meeting

For further information, please visit

<https://www.ratswd.de/en>