A Domain Data Protocol for survey data

Reiner Mauer and Jonas Recker, GESIS – Leibniz Institute for the Social Sciences

Workshop “Open Science and Sharing Research Data: Towards European Guidelines on RDM procedures”
Overview

1. Design principles: our ideas for the draft protocol
2. Examples
3. Challenges and open questions
Design principles: Scope

- Focus on survey research rather than the entire social sciences
- Suggestion: different protocols for different “styles of research” (e.g. quantitative and qualitative) or different types of data
Design principles: Structure

- Use of DCC DMP-Checklist headings

- Alternative: align with H2020 DMP template for “FAIR Data Management”
  - Advantage: brings FAIR principles to the foreground
  - Disadvantage: introduces potentially confusing redundancies. (E.g. FAIR requires the topic of metadata to be dealt with in different sections: interoperability, discovery, usability of data)
Design principles: Structure

- statements of requirements (“answers”) plus guidance (links + references to external sources with best practices/ specific aspects / details)
Design principles: Language

- “Confirmatory”: concrete description of requirements that researchers have to fulfill
- “We confirm that we are doing x, y, z” versus “We ensure that we will do everything that is necessary”
- “Comply or explain”
- if a project deviates from the protocol, it is expected to explain why and in which way
  ➔ important to meet the right level, otherwise the protocols will not deliver the expected efficiency gains for all parties involved.
Design principles: Baseline level

Meeting FAIR principles as minimum requirement - complying with the protocol means that data are considered FAIR

might be too ambitious as a standard level given the specific emphasis of the FAIR principles on enhancing the ability of machines to automatically find and use the data
6.3 Data formats

The project confirms that data will be processed and shared in a preferred or accepted format from one of the following lists:

- UKDS:
  https://www.ukdataservice.ac.uk/manage-data/format/recommended-formats
- DANS:

Guidance and further information

Examples

2.2 Documentation

The project will create the following documents during the research to ensure that the research findings are **reproducible** and that the research data can be understood and re-used. *(to do: specify by whom the data needs to be understood: researchers in the social sciences, in neighboring disciplines, of all disciplines...):*

- Codebook *(to do: define elements/generate template)*
- Method report *(to do: define elements/generate template)*
- Documentation of measurement instruments
- *

The project will archive and share these documents together with the data after the end of the funding period.

**Guidance and further information**

→ CESSDA ERIC (2017): Expert tour guide on Data Management: Chapter 2. Organise and document - Documentation and metadata. Available online:
Examples

6.1 Archiving

The project confirms that it will submit final versions of all research data and documentation required to replicate the (published) findings of the project to a recommended data archive, research data center or other suitable organization within xx months (specify acceptable embargo periods) after the end of the funding period. The selected organization guarantees that the data will be archived and remain available for at least xx years (specify minimum retention period).

In the case that personal data obtained by the project is made accessible to third parties (e.g. when data is published or archived), the project confirms that

- the consent form contains no phrasing that prevents the data from legally being shared;
- all protection measures guaranteed in the consent form are complied with.

Guidance and further information

→ CESSDA ERIC: Consortium of European Social Science Data Archives: https://www.cessda.eu/Consortium
Open questions and challenges

- FAIR as minimum level?
  - Does all data have to be FAIR? What is FAIR? What are accepted deviations from FAIR?
- Just one level or several?
  - a minimum level that everyone has to comply with
  - a “good practice” level
  - a high end/ gold standard level as maximum benchmark
- Community involvement in development, implementation and adoption
Open questions and challenges

- Striking the balance between being too specific and too general when describing the requirements
- Integrate Protocols in DMP tools?
  - e.g. Research Data Management Organiser, DMP online
Thank you!