Ingest: best practice in documentation

Brian Kleiner

FORS

SEEDS Workshop II

Ljubljana, 9. – 11. February 2016

seeds

South-Eastern European Data Services

Documentation in science - general

- The description of "how I came to my conclusions" is a fundamental basis of science and knowledge production

- Documentation is necessary: to be credible, valid, judged, and accepted by peers

Documentation and metadata

Documentation is any information that serves as a record of a research project and that renders data usable and meaningful.

- For social science research, documentation:
  - explains the study context;
  - explains the research process and how data were created, including instruments used;
  - describes the content and structure of data;
  - describes manipulations that have taken place

Metadata are defined as data providing information about one or more aspects of specific data

3

Documentation – in documents

Study-level documentation can be included in:

- final reports;

- technical reports;

- codebooks;

- questionnaires;

- interviewer instructions;

- interview topic guides;

- user guides;

- methodological papers;

- working papers or publications.

Study-level documentation

Good study-level data documentation includes information on:

- the context of data collection

- data collection methods

- data sources used and provenance of materials

- quality assurance procedures

- modifications made to data over time since their original creation

- for time series or longitudinal surveys, changes made to methodology, variable content, question text, variable labeling, measurements or sampling

- information on data confidentiality, access and use conditions, where applicable

5

Documenting quantitative data

Data documentation can be embedded in data, such as variable and code descriptions in datasets. Alternatively, information about data items can be recorded in a structured document.

Quantitative data-level documentation includes:

- names, labels and descriptions for variables

- values for response categories

- explanation of codes and classification schemes used

- codes of, and reasons for, missing values

- derived data created after collection, with code, algorithm or command file used

- weighting variables created and how they should be used

6

Documenting qualitative data

Qualitative data must be documented if they are to be used properly be the data producer, or reused, either by the original researcher (later), or by others. Many of the same elements of documentation apply for qualitative data, with some additions:

- study purpose and background;

- interview setting (participants, location, time of day, etc.);

- background of respondents;

- how data were anonymised

- data list describing cases, individuals or items studied, for example for logging qualitative interviews

Metadata and standards in the social sciences

The purpose of structured metadata is to allow for exchange of information about aspects and categories of research. The use of standards for metadata allows data in different places to be discovered and shared.

There are different existing metadata standards, such as:

- The Data Documentation Initiative (DDI)

- Dublin Core Metadata Initiative

- Metadata Encoding and Transmission Standard (METS)

- International Organization for Standarization (ISO)

- General International Standard Archival Description (ISAD(G))

8

CPC
Centre for Political Courage

Institute for Democracy and Mediation
IDM

ИНСТИТУТ ЗА
СОЦИОЛОШКИ И
ПОЛИТИЧКО
ПРАВНИ
ИСТРАЖУВАЊА

www.seedsproject.ch

FORS
explore.understand.share.

CeMI
CENTRE FOR MONITORING AND RESEARCH

adp

1958
COGNOSCERE EST MUTARE