Data Management Planning

OSP Awareness Session
October 30, 2019

Paul Bern, Syracuse University Libraries
Colin Elman, Political Science, QDR
Sebastian Karcher, QDR
Why be open about evidence-based inquiry?

• to achieve research transparency

• To enable secondary analysis, i.e. so another scholar can analyze the data

• for pedagogical purposes.
Our prescriptive methodologies all involve:

- extracting information from the physical and social world,
- analyzing the resulting data, and
- combining evidence and its analysis to reach a conclusion.

Examples of evidence-based inquiry in social science: ethnographic field work, a laboratory experiment, archive-based case study, or the statistical analysis of a large data set.
Elements of research transparency

Requires providing four types of information needed to evaluate an evidence-based claim.

- *Data citation* – what data were used and where are they?
- *Data access* – how can scholars obtain the data?
- *Production transparency* – if authors’ own data, *documentation* describing how the data were generated/collected.
- *Analytic transparency* – explain what steps were conducted with the data to arrive at the conclusion, and how the evidence and claims are connected.
Heuristic principle: data that were used to make the published claim

Data Sharing
Data made available for secondary analysis (aka “study dataset”)

Research Transparency
Data used to support a published evidence-based claim (aka “replication dataset”)
Virtues of research transparency

• Open science is more understandable, more complete, more evaluable, and more rigorous.

• Stronger results are more useful in developing solutions to substantive societal problems.

• Openness makes costly science more effective and efficient, thus delivering a substantial societal contribution.
Definitions

Research data management is caring for, facilitating access to, preserving and adding value to research data throughout its lifecycle.

Source: University of Edinburgh Information Services

A data management plan (DMP) helps researchers consider during the research design and planning stage, how the data will be managed during the research process itself and potentially shared afterwards with the wider research community.
Why manage research data well?

- Your data creation is likely to be expensive
- Your data underpin your published findings
- Good quality data = good quality research
- Protect your data from loss, destruction
- Compliance with ethical codes, data protection laws, journal requirements, funder policies
- To benefit your future self
Paul Bern: Data Management Plans
Types of Data Repositories

• Institutional repositories
  • E.g. SU Surface
  • Mainly designed for publications, though some take data

• Self-deposit repositories
  • E.g. Harvard Dataverse, Zenodo, OpenICPSR
  • No-cost for access/deposit
  • No or minimal curation

• Domain repositories
  • E.g. QDR, ICPSR, Protein Databank
  • Curation and assistance
  • List at [https://www.nature.com/sdata/policies/repositories](https://www.nature.com/sdata/policies/repositories)
DMP vs. IRB

- IRB: Required based on federal regulations
- DMP: Required by many funders, including NSF
- *Both require you to document data collection and sharing procedures*
- Critical to ensure DMP and IRB application align
DMP: Your Audiences

• Your funder
  • Main interest: funding impact, data sharing
• Grant reviewers
  • Main interest: your proposal, does the DMP support your proposal
• Your team
  • Co-authors, student workers, translators, transcribers, committee members
  • Main interest: data collection, does DMP help convey principles to team members
• Your future self
  • Main interest: does DMP help remind you of/prepare you for key decisions
What Funders Want

• Maximize impact
• Data sharing
• Use of suitable infrastructure
• DMP Tool
• Data repository
• NSF: Explicitly allows for DMP budget
Grant reviewers

- DMP is part of review package
- Panel *has to* comment on (i.e. at least skim) DMP

DMP = Additional 2 pages to showcase your competence
• DMP as reference document for team members
  • “How did we say we were going to store/organize...”

• Be specific! DMP can&should contain specifics about file/folder organizations, formats, etc.

• DMP as “living document” – keep your DMP updated with changes in plans
• DMP: Opportunity to plan key decisions ahead of time
• Check with others involved
  • IRB
  • IT
  • Data Repository
  • Local partner organizations
Stay in Touch

• qdr@syr.edu
• Twitter: @qdrepository
• Blog: https://qdr.syr.edu/qdr-blog
• Available for curation consultation for DMPs involving qualitative data