

What's Your Data Plan?



A data plan is a document that describes data and metadata that will be gathered in a study. It also includes details about the processes for preserving and sharing the data and indicates which procedures will be needed to access and use it.

A data plan helps the researcher map out which data and metadata to collect, and how to organize and store it, based on the ways the data will be used and shared in the future.

An external funder might require that you include a data plan as part of a proposal.

A data plan could begin by answering some of the questions on the right . . .

What are your data like?

- A. Make a list of the types of data you have or will have. Break it down by topic/project and type.
- B. What formats are your data in? Will you need to migrate them to new formats in the future?
- C. How much data do you (or will you) have?

How is the data organized?

- A. Are your data in spreadsheets? A database? A set of computer folders? A box on a shelf?
- B. Do you have metadata? Are they based on a standard?
- C. Which format(s) are your metadata in? XML tags embedded in data files? Separate text documents?
- D. Are your data structured using a data model, ontology, or vocabularies?

How will the data be preserved?

- A. Where will the data be stored? How will you back it up?
- B. How long will you need to preserve it? What resources will be required to retain it over this time period?
- C. Will some of the data be discarded? What criteria will be used to determine what is discarded?

How will the data be accessed?

- A. Who will access the data? Your research group? Your collaborators? Anyone in the world?
- B. What special methods will be used to ensure that sensitive components of data are kept private?
- C. What methods will be used to make the data or portions of it accessible, if required by a funding agency, publisher, etc.?
- D. Will it be deposited into a disciplinary or institutional repository? Made available as down-loadable files?

How will the data be used?

- A. Are any software or hardware tools needed to read or display it? How will these tools be made available?
- B. Are metadata, codebooks, or other forms of documentation needed to understand the data? How will they be provided?
- C. How should someone cite you and your data if they re-use it?