

The DCP Toolkit -- Summary of Interview Elements

Summary by Kristin Partlo of the "DCP Interviewer's Manual" of the Data Curation Profiles Toolkit, developed by Scott Brandt and Jake Carlson of Purdue University Libraries, the Institute of Museum and Library Services and the Distributed Data Curation Center. <http://datacurationprofiles.org/>

1. The data set (brief description)
2. The lifecycle of the dataset
 - a. Describe multiple stages of the data. For each, what are the number, size, and format of files?
3. Sharing
 - a. For each stage, would you share with: no one, your collaborators, within your institution, within your field, beyond your field, anyone? Timing of sharing?
 - b. What do you imagine the value to be to others? How might it be used?
 - c. Would you want to place any conditions on the use?
4. Access
 - a. Willingness to submit data to a repository? At what stages in the lifecycle? Embargo?
 - b. Prioritizing needs such as: ability to cite the data, requiring users to cite the data, mirror sites and restricted access.
5. Transfer of data / ingest into a repository
 - a. What preparatory actions would need to take place to make your data ready for ingest?
 - b. Prioritizing needs such as: ability to self-deposit, to automate submission via some trigger such as date or event, and the ability to batch upload.
6. Organization and Description of Data
 - a. Explain how the dataset is organized and how it has been described in documentation.
 - b. Is the organization and description sufficient for someone else to use it?
 - c. Prioritize need for ability to: make data available in multiple formats, apply standardized metadata from your field to the dataset.
7. Discovery
 - a. Prioritize ability for researchers within your own field to find the dataset, outside your field, general public, using Internet search engines.
8. Intellectual property
 - a. Who is the owner of the data? What are your funding sources?
 - b. Are there requirements from funders to: draft a data management plan, share your data with others, preserve your data beyond the life of funding? Any privacy or confidentiality concerns?
9. Tools
 - a. Software or hardware tools used in generating the data, utilizing the data.

- b. Prioritize need for ability to connect data to visualization tools, and for others to comment on or annotate the dataset.

10. Linking/Interoperability

- a. In the journals where you publish, are data or supplemental materials accepted for publication?
- b. Prioritize needs to connect data with publications, support the use of APIs, connect or merge your data with other datasets.

11. Measuring impact

- a. Prioritize needs for usage statistics, gather information about the people who use the datasets.
- b. Any other measurements or analytics you want to apply to your data?

12. Data management

- a. What are the primary ways you currently manage your data, including storage media and tools?
- b. Do you make backup copies? If so, how often? Do you take security measures to protect your data? If so, what? Prioritize need to enable version control.

13. Data preservation

- a. What are the most important parts to maintain or preserve over time?
- b. How long would your data have value for others if preserved?
- c. Prioritize needs for: ability to audit for structural integrity of files over time, migrate to new formats over time, secondary storage site for the data, secondary storage at a different geographic location, documentation of any and all changes made to the dataset over time.