Data Interview Protocol - Private Industry and Non-Academic Data Management Practices

Introduction

The goal of this interview is to gain a thorough understanding of how the interviewee's organization or organizational unit approaches data management tasks and problems. Topics of interest include:

- What workflows are used for data acquisition, including assigning metadata or other documentation to the data during acquisition;
- How data are preserved for reuse at a future date, and how decisions are made to keep or otherwise make use of collected data;
- How data are shared inside of the organization and with potential users outside of the organization.

Interview Questions

1) **Context**
   a. Please give an overall description of your work.
   b. What data do you generate in this work?
   c. What methods do you use to collect/create the data?
   d. What methods do you use to analyze data?

2) **Overview**
   Please describe the data flow for your organization. What types of data are generated or collected, how are they modified, used and/or analyzed internally, and how are they presented to users (internal and external)?

   Does the work you and/or your unit perform(s) have a more narrow scope? If so, can you describe the data flow for your unit and how it relates to the larger organization?

3) **Types of Data**
   This section will focus on the specific types of data generated in this research project including the type, collection methods, format(s), and any other information that is relevant to fully understanding the details of the data.
   a. Describe each data product this project will generate
      i. What methods (broadly) are used to generate or collect this data?
      ii. How many of these datasets are generated and how large are they?
      iii. In what general formats are the data (e.g. text, audio, video)?
      iv. Are there any special notes or comments about this data product that would be useful to record?
   b. Are there any overarching comments about the types of data you deal with that you think are relevant to understanding your work?

4) **Data Management**
   In this section we will discuss the data management practices of company or organization personnel in order to better understand how data are stored, backed up, and organized.
   a. Is there a formal plan for data management for your organization or group? If so, what is it?
   b. Who is responsible for managing the data?
   c. Do you have any internal or externally applied standards for file naming or organization?
   d. How are your data backed up?

5) **Data Documentation**
   a. How are your datasets documented (data dictionaries, code books, formal or informal metadata standards, etc.)?
   b. What are the major barriers to or problems with documentation of data in your organization?

6) **Data Sharing**
   Data sharing may be common within your organization and with consumers of your content but may sharing practices may differ in those contexts . In this section we will discuss the actual data sharing practices of your organization and how those practices change over time.
   a. How and when are data shared within your organization?
   b. How and when are data shared with consumers of your content?
   c. Are there, or should there be any restrictions on sharing any of the datasets you mentioned in the previous sections?
   d. How do you enforce restrictions on or track how your data is used?
   e. What are the major barriers to or problems with documentation of data in your organization?

7) **Please add any comments you might have about the above questions.**