Exploring Disciplinary Metadata Practices to Improve Georgia Tech Data Repository Services

Interview Protocol

Background information about the project

- We currently offer data archiving services using our institutional repository SMARTech.
- We are preparing to expand this service and build a data repository, because SMARTech was built for publications, and the Library has outgrown the technology.
- In the past, we have used a one-size-fits-all workflow for data deposits. However, given the specialized nature of many datasets, we are hoping to better tailor our approach.
- A critical component of data archiving (with an eye towards reuse) is the inclusion of appropriate metadata and documentation, to allow for someone to find the data, as well as understand those data well enough to use them in their own work.
- By better understanding what those forms of metadata and documentation are, we hope to enhance our support of diverse data deposits.
- In addition to asking questions about your experience working with data, specifically your experience documenting your data, we will ask you to review our current metadata forms, as well as a list of metadata items we’ve culled from your publications.

Ask the participant if they’ve reviewed the consent form, and provide them with a copy if necessary. Confirm that they consent to be a participant by signing the form.

Confirm that the participant is comfortable with being recorded.

Questions

**We’re making an assumption that because the interviewees self-selected and volunteered for an interview that they are generally open to archiving or sharing their data. If this proves to be an incorrect assumption, and the participant is skeptical of data archiving or data sharing, the questions can be reframed to focus the practices of their internal collaborations and research groups.**

1. What research data* do you create for your work/research?
2. How do you document your data or your work with the data (examples of documentation include lab notebooks, file naming conventions, README files, or change logs)?
3. Who typically creates the documentation?

Last update 11.12.14
4. What determines what types of documentation or metadata you collect? Are you following a community protocol, continuing the methods you were taught, or is it determined organically, as you recognize a need?

5. What information would someone need to know about your data in order to use them properly? This could be general concepts, like the methodology of the data collection, or specific pieces of information, like the orientation of the MRI scanner.

6. [Using the list compiled from analyzing the participant’s published literature] Which of these elements on this list would someone else need to know?

7. Do you use, or are you aware of, any tools to help with the creation and storage of your documentation and metadata?

8. Have you shared data with people outside your research group?
   a. If yes, what additional information did you provide and/or did they request?

9. Have you ever deposited data into a repository?
   a. If yes, what additional information did you include? Where there pieces of information that were required?

10. Do you use a disciplinary metadata schema to document your data? Are there any standards for collecting metadata or documentation that you use in your work?

11. Does your discipline use any standard file formats or have data standardization guidelines/best practices?

12. Have you ever used data that you yourself did not collect? This could include data collected by graduate students, collaborators, or researchers they have never before worked with.
   a. If yes, what additional information did you need in order to use those data?

13. If you were asked to fill out the two metadata templates (which were either reviewed ahead of time or are provided at this time), which fields would you fill out? Which fields are confusing? Which fields don’t need to be there? Which fields aren’t there that you think should be?

14. Do you have any other questions or concerns, or is there something you think we should have asked but didn’t?

*We have defined research data to be: Digital information structured by formal methodology for the purpose of creating new research or scholarship. May be in a variety of formats suitable for communication, interpretation, or processing.

Examples include:

- Observational data (e.g., sensor readings, survey instruments)
- Experimental data (e.g., lab equipment readings)
- Simulation data (e.g., climate models)
- Derived or compiled data (e.g., compiled databases, text or data mining)