# Reflection

# Yunzhao Wu

# Teachers College, Columbia University

# This week’s reading materials focus on data documentation. It means describing the data before data collection (Siiri Fuchs, & Mari Elisa Kuusniemi, 2018). By using metadata method, the goal of documentation is to fully explain what the data is during the work process. In this case, people who work together are easier to locate the files. An effective notation also can save time for teams to understand. In order to share resources more smoothly, the Common Education Data Standards (CEDS) project is a national collaborative initiative to create voluntary, common data standards to streamline the sharing, comparison and interpretation of data within and across institutions and sectors of P-20W for a key collection of education data elements (Common Education Data Standards (CEDS). n.d.). Corresponding examples are given to illustrate why common standards should be established. A readme file contains documentation about a data file and is intended to help ensure that the data can be interpreted correctly by yourself or others when sharing or publishing data at a later date (Cornell Research Data Management Service Group, n.d.). Readings also mentioned the software which can generate metadata automatically to save time.

After reading, I learned that data storage is like a book. The established standards and notes are just like catalogs, so people can find information faster. I want to know whether these standards and notes can be quickly searched for non-academic users. For example, a student wants to find out the frequency of use of mathematical formulas in a database. He may not know the technical terms. Then, can the student quickly understand the sentence of data storage to search?

The example I want to share is about my current work. The thesis I developed from readings is clear statements and standards makes working process efficiently in real-world situation. I’m working in a student organization, and one of my jobs is to create questionnaires to collect information. Whenever we plan to hold an event, I need to collect relevant information about the participants, such as name, email address, profession, etc. Then the information is generated as excel or csv file for other staff to browse. Excel is more readable because team members can see everyone's information very clearly. However, if I use tools such as R to analyze the data, the generated images need to be elaborated. Although it may be just some images instead of text, I think this can also reflect an important factor in file storage-creating annotations and explaining.

The other scenario is about data documentation. Before creating a shared working platform, the communication between teams is achieved through Asana and Wechat. Although we can communicate in time by these platforms, the shared files are difficult to re-access. For example, a staff member uploaded the audience information he collected. Everyone can see the file immediately when he uploads it. However, as the chat page keep refreshing, that file will be forgotten. At this time, people may ask the employee to resend the file again, which greatly reduces work efficiency. Shared work platforms such as Google Drive and Shimo allow everyone in the team can see the process of team's work and the shared files at all times. Even employees who are not in a department can quickly find the information they want through the shared platform.

In summary, this week’s reading makes me understand what data documentation is. A document-shared environment can enhance data accessibility. Standards for metadata are standardized formats that use particular terminology or ontology in the data definition (Siiri Fuchs, & Mari Elisa Kuusniemi, 2018). Thus, creating a metadata is crucial in improving understanding. Even if I so not use metadata method during work, but building detailed explanations has the same principle.

References

Siiri Fuchs, & Mari Elisa Kuusniemi. (2018). *Making a research project understandable—Guide for data documentation*. <https://doi.org/10.5281/zenodo.1914401>

Common Education Data Standards (CEDS). (n.d.). *Why CEDS?* Retrieved January 10, 2021, from [https://web.archive.org/web/20201019092710/https://ceds.ed.gov/pdf/why-ceds.pd](https://web.archive.org/web/20201019092710/https:/ceds.ed.gov/pdf/why-ceds.pdf)

Cornell Research Data Management Service Group. (n.d.). *Guide to writing “readme” style metadata*. Retrieved January 10, 2021, from <https://data.research.cornell.edu/content/readme>