Andrew Quinn

LIS 545

Winter 23

The National Student Loan Data System

Data and Metadata

DATA

I had difficulty separating data from metadata, however over the course of my higher education, I've found that metadata are what I would describe as bare details that are core to the subject matter. With my understanding of data, it was raw data that would evolve into information, whereas metadata was information requiring background knowledge to decipher. For this assignment I chose data.gov for their commitment to information that is open and accessible to all.

I decided to search for Federal loan forgiveness data on the data.gov database. Which returned 90 datasets and filtered the data into topics, topic categories (yielded no results), tags, formats (which was impressive to see how many tags and formats there were to choose from), organization types, organizations, publishers, and bureaus.

In particular, I looked at the data from the Department of Education and the Office of Student Aid because it recently had 737 views. The origins of the data are from The National Student Loan Data System.

The stakeholders are Americans or people who are seeking higher education in America and using government assistance. The datasets could reflect areas of weakness that could lead to improvement of policies. As well as research institutions, economic organizations and policy makers and the general public.

The National Student Loan Data System dataset has 17 data resources available. All additional resources are formatted in XLS and contain information such as federal school code list, quarterly activity, and a portfolio summary of federal student aid.

These additional datasets have restrictions and can be accessed or used under certain conditions. Those restrictions are related to how the metadata and data are structured by using their data schema, which requires them to input the data in their schema.

There is a specific software required to access the files which are plotly and CartoDB. I have never heard of nor have any experience with either of these softwares. Because I did not have the specific software needed to open it, I was unable to download any of the data.

METADATA

I was able to find the metadata of the National Student Loan Data System dataset by scrolling down. I found it easy to navigate through it after clicking on the download and links to understand how it

operated. The metadata tags and phrases were very helpful due to them linking to additional datasets that include that selected metadata.

The metadata is conveniently found at the bottom of the file page and available for download. The metadata seems comprehensive and easy to navigate to find additional resources but I'm unsure how the information provided would help the everyday user. For the metadata to be accepted the metadata must follow the Project Open Data Metadata Schema v1.1. Which also relates to the restriction to how the data and metadata be accessed or used under certain conditions.

I believe what would enrich the metadata would be the inclusion of demographics, socio-economic status, disbursement, repayment, deferment, delinquency, and closure. Most of the information cannot be understood to the general user, using simple and everyday language would make it accessible. I found the discovery portion adequate but a hurdle we face is the terminology and the presentation of the information. The next step would be to add the metadata I mentioned above to be included in the current data set.

To make the unfamiliar users able to use the data there must be relevant terms normal people would use in relation to the dataset to balance out the government and policy language. In terms of publications, I believe the downloads and additional resources are published datasets that came from the dataset but I cannot be found in literature databases or general web search engines because they're datasets. The web search I ran for my original dataset return results leading to the FAFSA website.

I found my understanding of data and metadata was not too far off from what is the reality of how this information is used and made to be accessible. Data are bare details that require no background knowledge whereas background knowledge is necessary to understand and navigate metadata. However to make metadata more accessible, common language is paramount to interpreting the information.