Abigail Reeb
LIS 545
Assignment 4- Data and Metadata Profile
January 23rd, 2024 (Edited February 19th, 2024)
Word Count: 914

DATA

 I chose a dataset from the 'gov' options called FDIC Failed Bank List. Briefly, this document is a spreadsheet showing the bank names, the city and state where the banks originated, a certification number, the name of the bank that acquired the former, the date where the acquisition closed, and a fund number. There are 568 rows of data in this document and dates that go back to late 2000. This data originates from the FDIC, was published by the Division of Insurance and Research, and was intended for public access and use as stated on the *data.gov* webpage.

 To view this data, one needs either Excel or a similar program that can open a .csv file. I was able to open it with LibreOffice Calc when I attempted to do so. I'm not very familiar with the permanence of different data formats, so I'd be interested in learning if .csv is seen as a format that will be long-lasting. Something that surprised me was that there is a broken link on the page where the download is available. The link that is supposed to lead to the dataset in an html format shows an error when clicked (when I tried again a couple of days later, the link was no longer broken, so I'm not sure what happened there.) There is no licensing information associated with this data, nor any usage restrictions. There is a link to a page on copyright restrictions of government documents, which is a useful resource for anyone not familiar with the topic.

 When one opens the page *data.gov* and navigates to the data catalog, they find that there are 295,003 datasets for perusal. Regarding ways to discover data, the catalog page has a search bar as well as a map where the user can search datasets by location within the United States, which is an interesting tool I hadn't anticipated seeing here. Under the map is a list of topics such as local government, climate, older adults, and energy that users can select to narrow down the list of datasets, as well as similar options to search for datasets related to different categories, file types, bureaus, and publishers.

 Additionally, there is quite a bit of information on the actual listings of dataset results to help the user determine if one would be useful. Each listing has the formats of the files that are contained, allowing the user to be aware if they would have to download additional software in order to access it. The listings also offer a brief description of the dataset, how many recent views it has, and whether the dataset is regarding a city, state, or federal government. The discovery system for this catalog is really comprehensive and more impressive than many others I've encountered. After using both vague and specific queries, I was able to manipulate the filters and still come up with search results that were relevant.

METADATA

The metadata is available in a separate .json document just below the original data on the download page. I was surprised at the size of the metadata when I saw it. For a document that has so many rows of values for seven columns, I'd have thought that the metadata would be longer than five rows of text. It isn't laid out as I'd seen .json formatted before– in shorter rows and indented to show the beginnings and endings of objects, so I'm wondering if I'm not opening the document correctly.

On the webpage where I accessed the metadata, beneath the link to download the .json document, there is a table embedded in the website itself that was much more interesting to me. Instead of five lines of text that are difficult to understand, here the fields and values are clearly presented. I can see that the metadata was created and last modified in 2020, that the organization responsible for maintaining the data and metadata is FDIC Public Data Feedback, and what the publisher hierarchy is in order to locate the file. This is also where one can find the harvest object and source IDs, the identifier, the schema version, and the bureau code, which may be specific to government data.

I can't say for sure that no publication has cited this dataset, but there's no place on the *data.gov* page that reflects it has been, as some articles have. This may be because data/datasets are treated differently from other research objects like articles and reports. There's also no suggested citation on the landing page, which doesn't mean it hasn't been cited, but is an interesting difference from some other datasets. I also searched Google Scholar for articles published after the creation of the dataset that I thought may have cited the data, but couldn't find any. This search was not exhaustive, but I spot-checked about twenty articles' lists of references and came up empty. It's possible that some of the references of the articles I looked at were documents that themselves cited the dataset, but couldn't determine that. There is a part of the sidebar on the *data.gov* page that has links to share the dataset on Twitter and Facebook, but no counter to show how many times users have done so, which I think is a shame. This part of the assignment made me realize that I don't have an idea how to find out if a document had been cited, so that was a fun adventure!

URL: https://catalog.data.gov/dataset/fdic-failed-bank-list