

Compiled Report

Bianca Phipps

LIS 545

Data and Metadata Profile

The data, found [here](#), are information on banks that have closed in the United States since October 1, 2000, that were acquired by the Federal Deposit Insurance Corporation (FDIC). The data includes the name of the bank, the city and state where the bank was located, the acquiring institution, the closing date of the bank, and two categories, Cert and Fund, whose meanings would be understood by a person in the field. (I am not an economics expert and based off minimal research couldn't find definitions.)

The data come from the FDIC, specifically the Division of Insurance and Research. According to the website page where the data were downloaded, the FDIC is often appointed receivers for these failed banks. Neither the website page where the data were downloaded nor the website where the data are displayed in an interactive table form contain additional documentation on how the data were collected. My assumption is through documentation of these failed banks, either through acquiring institutions or directly through the FDIC.

Key stakeholders for the data include the FDIC, who not only collected the data but are immediately impacted by the data, as receivers of the failed banks. The US government is another key stakeholder, both as the parent institution of the FDIC and as a ruling body over the country affected by the failed banks. Users of the data may include economists interested in exploring trends present within the data or examining what the data may mean for the future state of banks; researchers beyond the economic field who use the data to augment research that doesn't have an economic scope but is otherwise impacted or influenced by economics; or other banking institutions.

There is one data file: a .CSV spreadsheet. The file formats offered are this .CSV spreadsheet and technically HTML, as the data is available to view online [here](#), as mentioned

earlier in the write-up. The software necessary to open and analyze the data files are software compatible with the .CSV file, and a computer with connection to the internet. I was able to open the .CSV file on my Apple laptop with the Numbers app.

The data set comes with no usage restrictions and is declared meant for public use. There was no license information, either, only this statement: “If this work was prepared by an officer or employee of the United States government as part of that person's official duties it is considered a [U.S. Government Work](#)” (FDIC, 2020).

Regarding the metadata that comes with the data, there is a JSON file available to view, which includes properties like access level, distribution types, keywords, and the publisher. There is an additional metadata table on the data.gov webpage which includes additional information like resource type, metadata created date, metadata updated date, schema version, and publisher hierarchy. This metadata is not included in the .CSV file, and there is no additional metadata included in the .CSV file. The metadata provided on the data.gov webpage is comprehensive in the sense that there are quite a few fields with information, though these fields are concerned with the data set as a whole, and less about the data contained within the data set. There are only four keywords, which I found to be lacking. The metadata is structured according to the DCAT-US Schema v1.1, or the Project Open Data Metadata Schema (FDIC, 2020).

Considering enrichment, particularly looking at additional information that can improve discoverability and understandability, there are a few adjustments I believe could be made. The first would be to enhance the keywords used to describe the dataset. Currently, there are only four keywords: “assistance-transactions”, “banks”, “failures”, and “financial-institutions”. Keywords concerning the decades the data covered may add in discoverability and understandability. Definitions of each field would have been useful, too. As mentioned

previously, there were two fields I wasn't familiar with, and could find no documentation of what those fields meant or what their values meant. Additional information on how the data was collected and how often the list was updated (quarterly? the moment a bank is closed and acquired by a governing institution?) would be useful, too. On the data.gov website, it says the metadata was last updated November 12, 2020, but the most recent entry in the .CSV file was a bank that closed on November 3rd, 2023.

As for publications, there were none listed or provided with the data set. I performed a search on Google Scholar to see what articles had been published that referenced this data set. To do so, I searched for "FDIC Failed Bank List" in quotations. It did occur to me that because the title of the dataset was general, it would be harder to know if references were specifically to this dataset, but I persisted. There were 198 results in Google Scholar, and for the most part the articles I was able to access and skim did seem to be referencing this specific dataset. Because most citation searches are meant to discover articles, and not datasets, I wasn't sure if the method I was using was sound and satisfactory, but for the purposes of this assignment and in the time constraints, I felt satisfied with my results.

Works Cited

DCAT-US Schema v1.1 (Project Open Data Metadata Schema) | *resources.data.gov*. (n.d.).

<https://resources.data.gov/resources/dcat-us/>

FDIC. (2020, November 12). FDIC failed bank list. data.gov. Retrieved January 23, 2024, from <https://catalog.data.gov/dataset/fdic-failed-bank-list>

FDIC: Failed bank list. (n.d.).

<https://www.fdic.gov/resources/resolutions/bank-failures/failed-bank-list/index.html>

Repository Profile

Initially, I examined three repositories. The first was the [U.S. Bureau of Labor Statistics](#). While I was first drawn to this repository because of its description on the re3data.org site: “The U.S. Bureau of Labor Statistics collects, analyzes, and publishes reliable information on many aspects of the United States economy and society”. Ultimately, I couldn’t justify putting the data set of failed bank information into the BLS repository. While it *does* have to do with labor, as a failed bank would inherently mean unemployed bankers and other staff members, the data itself had no information regarding employment or labor statistics in general. Additionally, I couldn’t easily find a similar dataset within the repository, which made me believe it was outside the scope. The second was data.gov. Data.gov would have been an acceptable choice, as the dataset fell within its collection scope, but since I found the dataset there, it felt a little like cheating.

I chose the repository govinfo, found [here](#), because the dataset I chose was federally collected data and comes from a branch (albeit extended) of the United States government. According to their about page, “with the exception of materials classified for national security, all official Government publications, information, or information dissemination products paid for with Federal funds originating from agencies of the legislative, executive, and judicial branches of the U.S. Government are within scope for ingest into GPO’s system of online access. This includes any publications that are made accessible by the agency on their public website” (Agency Submission, 2023). This absolutely fit the dataset I examined: the FDIC published this data set as a part of its role within the U.S. government, and the data are accessible through their public website.

The repository has a defined collection scope. As noted above, the collection “provides free public access to official publications from all three branches of the Federal Government” (About us, n.d.). While the repository is closed to general submissions, all federal agencies are allowed (and encouraged) to submit their materials. According to the re3data.org profile on govinfo, they accept standard office documents, structured text, images, and audiovisual data. As the data are housed in a single .CSV file, I feel confident this fits under the category “standard office documents”. Potential submitters are guided to their National Account Managers, a role an expert (or simply member) of the federal field would recognize and understand. This leads me to believe that the repository provides human assistance for submitters, and may even have human assistance assigned to each branch. Because the repository works directly with government agencies, there is no public-facing information regarding what should be in the Submission Information Package.

As far as accessing the data housed within this repository, there is no login required to download data. There are a few ways for a person to access the information, with direct file download and database query being chief among them. There were three ways to search: a basic keyword search; an advanced search that offered customization like date ranges and collection and allowed searching of multiple indexes, like Sudoc Class Number, Title, and Government Author, to name a few; and citation search, allowing a user to work backwards from known information to reach the document.

The repository doesn’t declare it is displaying metadata using a specific metadata schema in a way that is immediately apparent to an end user (me). However, a majority of the data I accessed had a MODS file attached to it. It’s unclear whether the metadata displayed on the website is drawn from the MODS schema. Basic metadata for every record includes “Branch”

and “Category”, with varying information provided dependant on what the record is. Due to the wide variety in data taken in and preserved by the repository, I understand why there’s such variance in metadata display. Still, there are some discrepancies: one record, for example, had the field “Government Author,” indicating the branch the record came from, whereas other records included that information under the field “Publisher.”

To examine the DIP of the repository, I downloaded a few .zip files to note similarities between records. Where applicable, each DIP included three XML files: MODS, PREMIS, and DIPS, containing the metadata for the item, and then the item itself in the forms in which it was submitted to the repository. For example, one DIP I downloaded included only a PDF; one contained a PDF and HTML and XML files; and a third contained a PDF, an XML file, and thumbnails.

Works Cited

About us . (n.d.). Wwww.govinfo.gov; U.S. Government Publishing Office. Retrieved February 6, 2024, from <https://www.govinfo.gov/about>

Agency submission. (2023, April 28). GovInfo; U.S. Government Publishing Office.
<https://www.govinfo.gov/about/agency-submission>

re3data.org: govinfo; editing status 2021-11-05; re3data.org - Registry of Research Data Repositories. <http://doi.org/10.17616/R31NJN1I> last accessed: 2024-02-06

[re3data.org: U.S. Department of Labor, Bureau of Labor Statistics; editing status](http://doi.org/10.17616/R3WK7C)
[2024-01-30; re3data.org - Registry of Research Data Repositories.](http://doi.org/10.17616/R3WK7C)

<http://doi.org/10.17616/R3WK7C> last accessed: 2024-02-09

Recommended data citation

Federal Deposit Insurance Corporation. 2020. "FDIC Failed Bank List." Division of Insurance and Research.

<https://catalog.data.gov/dataset/fdic-failed-bank-list/resource/a8cfc40d-bf6d-4716-bba6-04fdbdf5f9c1> [Identifier. I'm having trouble finding a DOI for this particular set of data, and don't think I should just create one? If you have any guidance on how to proceed/where I might find a DOI for this object, I would love to know!]

Considerations for long-term preservation

Currently, the .CSV file format is not proprietary. Due to an increase in digital resources, there are no current concerns regarding the file format going obsolete. The hardware required to access the data are a computer with the ability to run the software required to open the file. The software that can open a .CSV file includes Microsoft Excel, Google Sheets, Apple's Numbers, or OpenOffice. Due to the small file size, if it would be necessary to migrate the data into a physical printed sheet of paper, printing the .CSV file and archiving it as a physical document would be feasible. If such an act were to occur, it would be our recommendation that the metadata also be preserved physically, alongside the data itself.

Copyright License

According to data.gov, the initial source for this data, there is no licensing information available. Because this work was prepared by an officer or employee of the United States government as part of that person's official duties, it is considered a [U.S. Government Work](#).

Human Subject Consideration Statement

This data does not contain personally identifiable data about people. Conjectures could be made about individuals who worked at the banks at the time of their closures, but would require more research and data beyond what is provided here. No steps were taken by me to anonymize or otherwise adjust the data for privacy or ethical reasons, as the data did not initially include information about individuals, only institutions.