Andrea Berg
LIS 545B
January 25, 2023
Term Project: Data and Metadata Profile

Data and Metadata Profile

I have chosen to curate a [dataset on Starbucks union elections](#) that has been uploaded to
the Kaggle repository. The data consist of information on store location, election dates, election
results, and whether or not the state in question is a Right to Work state. The data was originally
collected by journalists at More Perfect Union, who gathered information from union organizers
and solicited email tips on ongoing campaigns, maintaining an updated [database](#) on their
website. The context provided by the More Perfect Union coverage makes it clear that the data is
being used as pro-union evidence in support of the power and scope of Starbucks workers'
historic wave successful union campaigns. The key stakeholders in this case are Starbucks
workers, union organizers, and supporters of the labor movement, as well as the More Perfect
Union journalists who have compiled the data, Brandon Conrady who published the dataset to
Kaggle, and the 2491 people who have viewed it there. However, in another context, this same
data could be used for corporate or anti-union stakeholders to support claims that Starbucks
should crack down on union organizing.

The dataset is downloadable as a single CSV file with 15 columns and 259 rows. There
are no usage restrictions, the dataset has a CC0: Public Domain copyright license, and no special
software is required to open or analyze the CSV file. While the CSV file contains column
headings and entries for all cells, all of the metadata lives on the Kaggle page and is not included
in the downloadable file. The Kaggle page features a metadata section, which for this dataset
includes information on Collaborators (the person who posted the data to Kaggle), Coverage (the
time period covered by the data; in this case December 2021 – May 2022), Provenance (which

links to the More Perfect Union article), License (Public Domain), Expected Update Frequency (here listed as daily, with the last update 5 months ago), and Activity Overview (a posteriori metadata on how the data have been engaged with since being published to Kaggle). The Kaggle page also includes an About Dataset introduction describing the history of the Starbucks unionizing efforts, providing contextual and descriptive metadata to better understand the data. The data is displayed as a table in the repository, and each of the 15 column headings also includes metadata defining that field. For example, the column labeled "RTWState" comes with the description "Whether or not state is right to work." Finally, there are subject tags to connect this data to other thematically similar data in the areas of Business, Politics, Economics, and Statistical Analysis. However, all of this metadata, even the column heading definitions, is included only on the Kaggle page and not in the downloadable CSV file.

I would consider this metadata fairly comprehensive, particularly when you include the About Dataset introductory description and the Provenance link to the More Perfect Union article that gives deeper social context for understanding the data. The column heading descriptions provide another layer of clarification that would make the data understandable to those not familiar with labor movement terminology. The metadata is not presented in a particular standardized metadata structure, but most of it is organized under the Kaggle Metadata headings which do provide a user-friendly structure for identifying and browsing the metadata. One important improvement that I would make to this dataset would be to make the metadata included in the Kaggle repository page easily downloadable along with the CSV, so that the raw data doesn't get separated from the metadata during reuse. I would also add more subject tags to increase the visibility of the dataset for number of links between this data set and other related data sets, and include more relevant labels like "labor" and "union." Finally, I think it would add

value to the dataset to include metadata about the process by which the data was collected. This could inform future researchers working with the data or hoping to create a similar database for a different context. Despite the relatively high quality of the existing metadata, these improvements would make the dataset more findable, meaningful and useful to future researchers.

The data have already proved useful to journalists and labor researchers in a wide variety of contexts, as shown by the many articles and data visualizations that cite the original More Perfect Union database from which this Kaggle dataset was pulled. The data forming the Kaggle dataset is limited to the time period of December 2021 to May 2022, while the database maintained by More Perfect Union has been updated as recently as January 21, 2023 as union elections continue to take place. Therefore, although I was not able to find any citations of the specific Kaggle dataset (using several Google keyword searches), the More Perfect Union Database appears to be a major authoritative source for data on Starbucks union elections. A Google search for "Starbucks union elections More Perfect Union" returned articles and visualizations from [Union Election Data](), [People's Policy Project](), [Common Dreams,]() and [the New York State Senate](), all citing the More Perfect Union database. There is clearly a high level of interest in this data, and I believe the metadata improvements proposed above would further increase the visibility of this data and provide even more value for labor researchers and journalists.

References:

Bruenig, M. (2022, November 8). *Starbucks Union is nearing 7,000 workers*. People's Policy
    Project. Retrieved January 25, 2023, from
    https://www.peoplespolicyproject.org/2022/11/08/starbucks-union-is-nearing-7000-
    workers/

Conrady, B. (Ed.). (2022, August 13). *Starbucks Union election results*. Kaggle. Retrieved
    January 25, 2023, from https://www.kaggle.com/datasets/brandonconrady/starbucks-union-
    election-results

Johnson, J. (2022, July 23). *'historic moment in American Labor': Starbucks Workers Notch
    200th Union win*. Common Dreams. Retrieved January 25, 2023, from
    https://www.commondreams.org/news/2022/07/23/historic-moment-american-labor-
    starbucks-workers-notch-200th-union-win

Mayernik, M.S. (2020). Metadata. In B. Hjørland & C. Gnoli (Eds), *Encyclopedia of Knowledge
    Organization*. International Society for Knowledge
    Organization. https://www.isko.org/cyclo/metadata

More Perfect Union. (2023, January 25). *Map: Where are Starbucks Workers Unionizing?* More
    Perfect Union. Retrieved January 25, 2023, from https://perfectunion.us/map-where-are-
    starbucks-workers-unionizing/

Ocasio, V. (2022, September 8). *Wantagh Starbucks Workers vote 20-7 for union*. NY State
    Senate. Retrieved January 25, 2023, from https://www.nysenate.gov/newsroom/in-the-
    news/anna-m-kaplan/wantagh-starbucks-workers-vote-20-7-union

Reuning, K. (2023, January 25). *Current Starbucks Statistics* . Union Election Data. Retrieved
    January 25, 2023, from https://unionelections.org/data/starbucks/