Final Project: First Draft

Karen Butler-Clary

LIS 4220 – Winter 2021

Dr. Matthew Mayernik

February 25, 2021

Data and Metadata Profile

The dataset selected for curation is the Museum Data Files from the Institute for Museum and Library Services (IMLS) which can be found on the imls.gov website (*Museum Data Files FY 2018*, 2018). It was most recently compiled in 2018 to “provide information about museums and related organizations in the United States” (*Museum Data Files*, 2021). The dataset includes information on approximately 33,000 institutions across the United States and the District of Columbia (*Data Catalog*, 2021). IMLS draws on many sources to pull together this dataset from other available sources. The dataset is compiled with the intention that it will be used to show the impact of museums on the communities they serve. However, the dataset could also be used for more practical information such as tracking where institutions are located, the discipline of the institutions and their IRS 990 revenue information.

Given the depth and breadth of the information, there are numerous stakeholders. Stakeholders include city planners and policy makers looking to determine how they could use community resources. Individual institutions could use the data to determine how their organization compares to organizations with a similar discipline, budget, staff size, etcetera. Employees could use the information to argue for salary increases or better benefits as comparable institutions provide. There are innumerable ways to utilize the dataset and this is by no means an exhaustive list.

The Museum Data Files is a set of three CSV Data files, each with one spreadsheet consisting of 50 variables per institution. Variables contain information on every institution in the United States that met the criteria for being a museum and included location information, reported budget and revenue, and discipline. The criteria defined by IMLS for inclusion in the dataset require an institution be: non-profit or government run, a permanent institution for essentially educational or aesthetic purposes, owns or uses tangible or intangible objects, cares for these objects, and exhibits those objects regularly to the general public through a facility it owns or operates (Frehill & Pelczar, 2018). No proprietary software is necessary to open and/or analyze the data files and no usage restrictions were found.

Also available with the dataset is a File Documentation and User’s Guide. The guide consists of several tables and appendices explaining the data. The descriptive metadata provided includes a key the numerous abbreviations and codes within the dataset (Frehill & Pelczar, 2018). Structural metadata is also supplied to describe how the three separate CSV files are arranged. Lastly the guide contains administrative metadata information such as usage rights, and when and how the dataset was created. The metadata covered within the File Documentation and User’s Guide is quite comprehensive, but it is also absolutely necessary for understanding the dataset. While some data elements are self-explanatory other abbreviations and codes require a key to extract their meaning. There is not sufficient metadata within to data files to fully utilize the dataset without the File Documentation and User’s Guide.

The dataset incorporates some metadata standards for particular elements but more so as a means of controlled vocabulary. The North American Industry Classification System (NAICS), Standard Industrial Classification (SIC) codes, and the Federal Information Processing Standards (FIPS) State and County Codes, all provide consistency among the elements within the set and other federal government reports. However, the dataset does not incorporate a particular schema, such as Dublin Core, to identify elements or make the data consistent with other datasets.

Applying a metadata schema could prove to make the dataset more useable and discoverable. There are elements in the dataset unique to this study but many could be plugged into Dublin Core or a similar type metadata standard with ease. DCMI could be utilized though it may need some adaption since this dataset contains information about organizations rather than objects. Expanding the codes and abbreviations would also be crucial to increase discoverability in a repository environment and necessary for an individual unfamiliar with the data to utilize the dataset.

Though no publications could be found yet, the dataset is being used in the IMLS study *Understanding the Social Wellbeing Impacts of the Nation’s Libraries and Museums,* which focuses on understanding “the conditions under which museums and libraries contribute to social quality of life, economic prosperity and other aspects of wellbeing” (*Museum Data Files*, 2021). However, given the dataset has been available for download for a few years, it seems very unlikely it has never been utilized in a publication even though none could be found.

Repository Profile

The repository selected for placement of the Museum Data Files from the Institute for Museum and Library Services is the World Values Survey WVS Data Archive (*World Values Survey*, 2020). The Museum Data Files are currently, and logically, housed within the IMLS Data Collection repository (cite). However, this dataset’s potential use extends beyond the reach of the IMLS Data Collection repository and would benefit from additional placement to reach new audiences. Placing this dataset in a repository with a different but complementary audience will allow it to contribute to new areas of research. The World Values Survey was selected because it does not currently contain any data similar to the Museum Data Files but it is believed the Museum Data Files would contribute a new perspective and research angle to the ongoing work of the Survey.

The World Values Survey (WVS) is an international research program devoted to the scientific and academic study of social, political, economic, religious and cultural values of people in the world. The project’s goal is to assess which impact values stability or change over time has on the social, political and economic development of countries and societies (*Who We Are*, 2020).

The Museum Data Files provide a cultural values map to the United States. When overlaid with political and economic maps, one could research how the presence of cultural institutions affects the economic and political leanings of a geographic area. While the Data Files can only provide this information for the United States, similar studies have been done internationally and could be added to the WVS Data Archive as well to complete this perspective for research.

The WVS Data Archive repository is not open to general data submissions but rather is the repository for all of the data collected from the questionnaires developed by social scientists for the organization. Since 1981, the World Values Survey Association has been collecting data from social surveys in 5 year cycles from 120 countries to create the “largest non-commercial academic social survey program” (*Who We Are*, 2020). The repository serves a very specific purpose with a very defined collection scope. However, the data contained within the Museum Data Files fits well with the WSV “project’s overall aim is to analyze people’s values, beliefs and norms in a comparative cross-national and over-time perspective” (Who We Are, 2020). Like the WSV survey, the Museum Data Files have been conducted at multiple time intervals complementing the over-time perspective of the WSV.

The World Values Survey provides access to seven different waves of the survey spanning the last 40 years as well as a 2017 joint report with the European Values Survey, regional WVS surveys, and a Times Series Data set that shows how the survey results have changed over time (*Data & Documentation*, 2020). All of the datasets are available for download on the World Values Survey Website as well as a quick-reference guide for using and understanding the datasets. No login is required to download any of the materials, though the user must register, at no cost, providing some very basic demographic information and agree to the WVS non-redistribution data use license to access the .rds data files. The reference guide, questionnaire, and documentation for the datasets is available in PDF form. The datasets are provided in multiple formats, SPSS, Stat, ASCIL text and Excel, and the recommended citations for the sets are also provided.

In addition to file download, an Online Data Analysis tool is available as an additional access mechanism provided for the datasets (*Online Data Analysis*, 2020). To use the Online Data Analysis tool no registration is required. The online tool allows users to view data in the visual forms of maps and various graphs. The data can be sorted by dataset, country, survey questions and specific responses and manipulated within the categories. The tool is easy to use and provided a PDF or Excel file download of the selected data analysis with no registration required.

The WVS Data Archive online tool and available downloadable files make for a simple and easy Dissemination Information Package (DIP) delivered online. The online tool in particular allows users to customize the DIP/s to the exact information desired. However, no specific metadata standard was discernable from the information viewed. The surveys do have clearly established, consistent Likert scales for answering survey questions. The database interface is a very well-done website that encourages use and reuse of the information. Overall, the Museum Data Files would be a good addition to the WVS Data Archive.

Long-term Preservation Considerations

Fortunately, this dataset is provided in multiple file formats which allows more opportunities for long-term preservation. It would be optimal to *Data File Documentation* for the IMLS Museum Data Files to have a PDF/A preservation file in addition to the PDF file for access and use. All of the datafiles should also have both XML and CSV formats available allowing for extensive use in a non-proprietary file format.

Copyright and Recommended data citations

The IMLS Museum Data Files are the result of a publicly funded project. Therefore, unless it is otherwise specifically noted, all the information contained within the IMLS datasets and the *Data File Documentation* are in the public domain but citation of the source is required with use (Frehill & Pelczar, 2018). The information contained within the World Values Survey datasets is the result of an internationally grant funded project that is also free and available for use to anyone but does also require citation for use. Citations for the datasets are available where the datasets are downloaded (*Data & Documentation*, 2020). For information accessed from the GitHub repository, the suggested citation is as follows (Butler-Clary, 2021):

Butler-Clary, K. (2021). *Adding IMLS Museum Data Files to WVS* [Personal]. Repository Term Project. https://github.com/kbutlerclary/repository.term.project

Human Subject Considerations

All data included in both the IMLS Museum Data Files and all of the datasets that constitute the World Values Survey have no personally identifiable information. Personally identifiable information was not collected for the IMLS Museum Data Files and therefore no considerations for human subjects need to be applied. The World Values Survey datasets are an internationally compilation of the work of many different social scientist and research groups. Before data can be submitted to the WVS by those participating groups, the data must anonymized prior to acceptance. Only anonymous data is available for use from the WVS and there is no way to trace survey respondents (*Fieldwork and Sampling*, 2020).

References

Butler-Clary, K. (2021). *Adding IMLS Museum Data Files to WVS* [Personal]. Repository Term Project. https://github.com/kbutlerclary/repository.term.project

*Data & Documentation*. (2020). [Organization]. World Values Survey. https://www.worldvaluessurvey.org/WVSContents.jsp

*Data Catalog*. (2021). Imls.Gov. https://www.imls.gov/research-tools/data-collection

*Fieldwork and Sampling*. (2020). [Organization]. World Values Survey. https://www.worldvaluessurvey.org/WVSContents.jsp?CMSID=FieldworkSampling&CMSID=FieldworkSampling#:~:text=WVS%20requires%20a%20complete%20explanation,by%20the%20EC%20in%20writing.&text=All%20deposited%20data%20has%20been,means%20to%20trace%20the%20respondents.

Frehill, L. M., & Pelczar, M. (2018). *Data File Documentation: Museum Data Files, FY 2018 Release.* Institute of Museum and Library Services.

*Museum Data Files*. (2021). Imls.Gov. https://www.imls.gov/research-evaluation/data-collection/museum-data-files

*Museum Data Files FY 2018*. (2018). Institute for Museum and Library Services. https://www.imls.gov/research-evaluation/data-collection/museum-data-files

*Online Data Analysis*. (2020). [Organization]. World Values Survey. https://www.worldvaluessurvey.org/WVSOnline.jsp

*Who We Are*. (2020). [Organization]. World Values Survey. http://www.worldvaluessurvey.org/WVSContents.jsp

*World Values Survey*. (2020). [Organization]. World Values Survey. https://www.worldvaluessurvey.org/wvs.jsp