Term Project: Hidden Barriers Data Set: Final Report

Grace Adler

University of Washington

LIS 545 B

Term Project: Final Report

URL to data set: Hidden Barriers: [The Experience of Academic Librarians and Archivists with Invisible Illnesses and/or Disabilities Data Set](https://figshare.com/articles/dataset/Hidden_Barriers_The_Experience_of_Academic_Librarians_and_Archivists_with_Invisible_Illnesses_and_or_Disabilities_Dataset/17161328?file=31767911)

URL to GitHub repository: <https://github.com/graceadler64/LIS545-Invisible-Illnesses-and-or-Disabilities>

**Recommended Citation**

Quirin Manwiller, K., Peter , S., Anderson, A., & Crozier , H. (2021, December 13). Hidden

barriers: The experience of academic librarians and archivists with invisible illnesses and/or disabilities dataset. GitHub. <https://doi.org/10.15786/17161328.v1>

**Digital Preservation Plans**

Both .csv and .txt files are popular file formats readable by software from multiple major corporations, like Google and Microsoft. Even if someone tried to replace this file format with their own company’s innovations, it would need to outpace the wide usage of .csv and .txt to be successful. While this lowers the risk of the file types becoming obsolete in the immediate future, data curators should do an annual review of news in the file formatting world to ensure corporations haven’t proposed a successful replacement file type. Data curators should also take this as an annual reminder to check the fixity of each file, so that bit rot is avoided.

**Copyright License**

CC BY-NC-ND 4.0  
 According to Creative Commons (2024), this means that users are allowed to share the data set as long as the researchers are credited, the purpose is non-commercial, and the data is not being altered substantially.

**Human Subject Considerations**

The Hidden Barriers authors (2023) note in their own ReadMe.txt that their data set has data from open-ended questions that involve personal experiences, but potentially identifiable raw data has “been removed to protect participant anonymity.”

Data and Metadata Profile

**Data Overview**

Quirin Manwiller, Peter, Anderson, and Crozier (2023) set out to identify hidden barriers faced by “MLIS-holding” professionals with disabilities or illnesses that may not be consistently observable, colloquially known as an invisible illness. Their data set includes quantitative data, like demographic information, qualitative data measured through quantitative values, like close-ended and Likert scale questions, and qualitative open-ended responses (Pragadeesh, 2021).

**Data Origins**

The data comes from an anonymized thirty-question survey conducted as part of research at West Chester University and the University of Wyoming. Three hundred and fifty-five respondents, geographically located in the United States, responded. However, around 60 were disqualified due to either not working in the library/archive field or not having an invisible disability (Quirin Manwiller et al., 2023).

**Key Stakeholders**

Key stakeholders for this data include library professionals at all levels of knowledge of invisible illness, position in their organization, and personal disability. A disabled individual could use this data to advocate for better accommodations, and similarly, a library administrator could use this data to make relevant budgetary, accessibility, and policy delicious. Researchers in other fields could cite this data to propose research into their own professional community’s experiences with invisible illness. West Chester University and the University of Wyoming also stand to benefit from this data as its citations could raise prestige, and the researchers themselves would share in that benefit.

**Data Files**

There are two files included, one is a CSV file that contains the anonymized responses of the 355 survey respondents, and the other is a TXT file that includes both the survey questions and relevant metadata. The CSV file does include the survey questions in the second row, but the default columns are quite thin. This means users either spend meticulous time expanding those columns or be lucky and stumble upon a more readable version in the TXT file.

**Software Requirements**

CSV files can be opened by “any text editor or spreadsheet program,” but most users would be using Microsoft Excel or Google Sheets (Carnes, 2019). TXT files can be opened by “all text editors” since there is “no special formatting” (Text File Format, n.d.). Specific examples include Microsoft Word, the Windows Notes app, Google Docs, and so on. The metadata states that the data was analyzed using Qualtrics, which is an online survey tool that helps users collect, analyze, and distribute survey responses (Quirin Manwiller et al., 2023).

**Usage Restrictions**

This data set has the following license: CC BY-NC-ND 4.0. According to Creative Commons (2024), this means that users are allowed to share the data set as long as the researchers are credited, the purpose is non-commercial, and the data is not being altered substantially. The researchers also included a DOI and recommended citations for easier, stable references.

**Metadata Content and Comprehensiveness**

The metadata covers the following: a title, author, abstract, keywords, discipline, geographic research location, license, other sharing/access rules, recommended citations, data overview, methodological information, variable counts, and so on. Needless to say, that is fairly comprehensive, but several key pieces of metadata are inaccurate or missing, such as not adding a DOI in the provided spot for a DOI to go. They also say there is not a published version of this data, which might have been accurate when added to FigShare in 2021 but is not true as of their 2023 publication discussed below.

**Metadata Location**

The metadata is exclusively in a separate TXT file, the data set itself has only questions and responses.

**Metadata Standard**

No metadata standard is referenced in the TXT file, so I referenced visuals of Dublin Core, MODS, CWM, and so on and did not locate metadata that looked similar.

**Potential Improvements and Data Set Publication**

The researchers should have included the survey questions as a separate file rather than including them at the bottom of the TXT file. Survey questions are data in their own right, as they are variables to the research being done. Including them in a separate file, even as an appendix like they did in their 2023 publication, would have made it easier for users to locate them (Quirin Manwiller et al., 2023).

There are also easy updates that are missing from the metadata, such as copying in the DOI. It’s easy to assume that when writing the metadata the DOI hadn’t been established yet, but it should have been updated upon publishing.

The metadata also references making graphs through the software Qualtrics, but these graphs are not included in the files uploaded to FigShare (Quirin Manwiller et al., 2023). The graphs are included in the [journal publication of this work at the College and Research Libraries](https://crl.acrl.org/index.php/crl/article/view/26024/33938#:~:text=Respondents%20reported%20barriers%20including%20professional,and%20disability%20in%20the%20profession.). The graphs are a much faster way to analyze the data than to start from scratch with the raw data without the observations that the users have already made. Ideally, these graphs, as well as a full PDF of the final published version, would be included in the FigShare files. Bringing it back to the metadata itself, a link to this publication should have been included in the metadata for greater accessibility and discovery.

The 2023 publication also includes [a Dublin Core metadata file](https://crl.acrl.org/index.php/crl/rt/metadata/26024/33938), which would be excellent to include in the FigShare files. For this metadata, the metadata schema could not be more clear as it is noted explicitly (Quirin Manwiller et al., 2023).

**Listed Publications**

There are not any publications listed or provided with the dataset. The “Read Me” file includes a space to add publications but it has not been updated since the data set was uploaded to FigShare.

**References to Dataset**

To search for citations of this data, I used a combination of Google Scholar and UW Library’s search engine. While I found similar research studies on disability in the library profession, a scouring of their references did not include this specific data set. According to the 2023 publication of the research, the article has been viewed roughly 2,463 times since its publishing, so it is not completely obscure in the field (Quirin Manwiller et al., 2023).

Term Project: Repository Profile

URL to data set: Hidden Barriers: [The Experience of Academic Librarians and Archivists with Invisible Illnesses and/or Disabilities Data Set](https://figshare.com/articles/dataset/Hidden_Barriers_The_Experience_of_Academic_Librarians_and_Archivists_with_Invisible_Illnesses_and_or_Disabilities_Dataset/17161328?file=31767911)

URL to repository: [ICPSR](https://www.icpsr.umich.edu/web/pages/index.html)

**Repository Choice**

I chose ICPSR (Inter-university Consortium for Political and Social Research). Although the study itself focused on professionals in the MLIS discipline, it deals directly with discrimination in the workplace and how disability affects individuals (Quirin Manwiller & Peter, 2021). This feels aligned with ICPSR’s focus on the social and behavioral sciences, and the data set could even qualify for specialized data curation in one of their twenty-one special collections, like the Archive of Data on Disability to Enable Policy (ADDEP), Health and Medical Care Archive (HMCA), Data Sharing for Demographic Research (DSDR), or the Resource Center for Minority Data (RCMD) (Data Collections at ICPSR, n.d.). They even include “Discrimination” as a suggested search on their homepage, so the data set is well within the scope of ICPSR’s research scope (ICPSR, n.d.).

The Hidden Barriers data set was created and analyzed by researchers with various university affiliations, such as West Chester University and the University of Wyoming, and all but one of the listed universities is a member of the ICPSR (List of Member Institutions and Subscribers, n.d.). This means that, in addition to the extensive deposit guides, general data education, and support ICPSR provides to non-members, the researchers are allowed to submit to the membership archive without a fee (Start Sharing Data, 2018). The researchers could request funds from their member institutions more easily if they need data curation beyond what ICPSR typically offers (the cost for which varies depending on the data size and need), and ICPSR will share their data with members and non-members at no additional cost to them (Start Sharing Data, 2018).

**Repository Limitations and Restrictions**

ICPSR does not have limits on data type or file formats, though they do have recommendations for optimal reuse and storage (Depositing Data with ICPSR, n.d.). They do not explicitly forbid data sets outside the social and political sciences, but the financial information below might make it suboptimal for research outside the scope of its sponsored collections and membership institutions (Depositing Data with ICPSR, n.d.).

ICPSR states that it is “usually” free of charge, like for submissions that fit into the scope of a “sponsored archive (i.e., National Archive of Criminal Justice Data),” but non-members hoping to submit to the member-specific archive would need to either seek sponsorship or pay “$550 per study” (Depositing Data with ICPSR, n.d.). This financial detail is important to note because the level of curation that the data receives is explicitly tied to whether you are a member, if your data set fits into the scope of a sponsored collection, and so on. A non-member data set on a subject completely irrelevant to ICPSR’s specialized social and political science fields might be able to be submitted, but it is less likely to get sponsored curation and has fewer Open Access options (What Are the Differences between ICPSR’s Deposit Options, n.d.).

ICPSR explicitly accepts data submissions from “member and non-member institutions” and has guidelines to deposit both public and restricted-use submissions (Start Sharing Data, 2018). They even include a [helpful flowchart](https://www.icpsr.umich.edu/files/datamanagement/deposit-services-d5.pdf) to help potential submitters understand what they accept, which includes unique inclusions like “citizen scientists” and complex projects that require in-depth work to be ready for a Submission Information Package (SIP).

G**uidance for the Submission Information Package (SIP)**

ICPSR has an abundance of submission guidance.

They request, but do not require, that quantitative data is submitted as “SAS, SPSS, Stata,” or ASCII files with variables that are accompanied by a set of “exhaustive, mutually-exclusive codes” (Start Sharing Data, 2018). They request, but do not require, that qualitative data is “well-documented,” organized by nested file folders, and in the following file formats: .txt, .rtf, .pdf, .doc, or .nvp (“Guide for Sharing Qualitative Data at ICPSR,” 2023).

Other main requirements are that the SIP includes all relevant data files and documentation files, that it has a study-level description, and that all personally-identifiable information about human subjects meets industry standards for informed consent (Start Sharing Data, 2018).

ICPSR also provides a sixty two page guide with helpful topics to potential submitters including recommended elements, example data management plans, metadata best practices, respondent confidentiality, and so on (Guide to Social Science Data Preparation and Archiving, n.d.). Their recommended, but not required, elements include: a data description, review of existing data sets, expected formats the data will be in during the submission/distribution/preservation stages, structured metadata, and so on (Guide to Social Science Data Preparation and Archiving, n.d.).

**Human Assistance and Submission Consulting**

A help email is offered ([ICPSR-help@umich.edu](mailto:ICPSR-help@umich.edu)) but it is not submission specific. There is also a listed phone number (734-647-2200) but it is also not submission specific. They also mention that members could have a “local Representative who can help” depending on their partner institution (Help and User Guide, n.d.). Finally, there is a “[User Support Request Form](https://cms.icpsr.umich.edu/form/user-support-request-form?utm_source=website&utm_medium=button&utm_campaign=USRF&utm_id=USRF)” but it is not submission specific. From this combination of resources, help pages, and contact points, it seems like most direct human assistance is done through older digital communications like email and telephone, rather than over a video call, live chat, or in-person. The important exception being, of course, if you are a member and happen to have a local representative who is knowledgeable about preparing a SIP.

**Metadata Submission Requirements and Standards**

They do not require a specific structure, but recommend the metadata is in structured XML and list the following as “important metadata elements”: principal investigator(s), title, funding sources, data collector, project description, units of analysis (what or who is being studied), variables, related publications, technical file information, data collection instruments, index, coding information, and “Readme.txt” files (Guide to Social Science Data Preparation and Archiving, n.d.).

**Login Requirements**

After experimentation, a login is required to download complete data files from Open ICPSR. Both a login and membership are required to access member-only data sets. Without a login, one might be able to download a limited number of files, like metadata or Readme.txt, but the raw data itself will likely be missing. This was tested with both Open ICPSR data and member funded data, but there is a lot of variation between what individual data sets provide. Even with membership, there are some studies with restricted data ([example](https://www.icpsr.umich.edu/web/ICPSR/studies/38275)) but ICPSR provides guidance on how to bypass the restriction with general tips like submitting your project summary, academic affiliations, confidentiality pledges from relevant staff, and so on (Accessing Restricted Data at ICPSR, n.d.).

**Access Mechanisms**

ICPSR provides direct file download so users can analyze them on their own computers, but some data sets can be analyzed through their site itself (Access and Dissemination, n.d.). You do have to log in to access in-site analysis, and it is provided through Survey Documentation Analysis (SDA) (Analyze Data Online with SDA, n.d.). When previewing a data file that can be analyzed online, you can browse and search variables. Some studies allow you to “explore the data” by adding or removing potential variables, generating tables based on selected variables, and so on (Explore Data, n.d.).

**Specific Metadata Standard**

The metadata is available in multiple formats depending on the specific data set. From some exploratory searching, some offer just Dublin Core and DDI 2.5, while others include up to five exportable metadata standards for the same information. For its core projects, ICPSR states its metadata is “based on the Data Documentation Initiative (DDI)...which ICPSR helped establish and now leads” (Metadata, n.d.).

**Dissemination Information Package (DIP)**

The content of what actually disseminates to users from ICPSR depends on both what the user chooses, and how many files were in the original SIP from the researchers. The DIP could include just the data set by itself, a codebook, exportable metadata, various preexisting analysis, and/or other accompanying materials (Access and Dissemination, n.d.). ICPSR also assigns a Digital Object Identifier (DOI) and recommended citation to each study to allow for easier navigation to find the DIP from external sites (Additional ICPSR Services, n.d.).

References

*Access and dissemination.* (n.d.). ICPSR; University of Michigan. Retrieved January 31, 2024, <https://www.icpsr.umich.edu/web/pages/datamanagement/lifecycle/access.html>

*Accessing restricted data at ICPSR.* (n.d.). ICPSR; University of Michigan. Retrieved January 30, 2024, from <https://www.icpsr.umich.edu/web/pages/ICPSR/access/restricted/>

*Additional ICPSR services*. (n.d.). ICPSR; University of Michigan. Retrieved January 31, 2024, https://www.icpsr.umich.edu/web/pages/datamanagement/lifecycle/access.html

Analyze Data Online with SDA. (n.d.). ICPSR; University of Michigan. Retrieved January 31, 2024, from <https://www.icpsr.umich.edu/web/pages/RCMD/tools/sda.html>

Carnes, B. (2019, November 1). What is a CSV file and how to open the CSV file format. Free Code Camp. <https://www.freecodecamp.org/news/what-is-a-csv-file-and-how-to-open-the-csv-file-format/>

CC BY-NC-ND 4.0 deed | attribution-noncommercial-noderivs 4.0 international | creative commons. (2024). Creative Commons. <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en>

*Depositing data with ICPSR.* (n.d.). ICPSR; University of Michigan. Retrieved January 28, 2024, from <https://www.icpsr.umich.edu/files/datamanagement/deposit-services-d5.pdf>

*Explore Data*. (n.d.). ICPSR; University of Michigan. Retrieved January 31, 2024, from <https://archive.icpsr.umich.edu/view/studies/34945/explore-data>

*Guide for sharing qualitative data at ICPSR*. (2023). ICPSR; University of Michigan. <https://deepblue.lib.umich.edu/bitstream/handle/2027.42/191150/Guide%20for%20Sharing%20Qualitative%20Data%20at%20ICPSR%20V2.pdf?sequence=3&isAllowed=y>

*Guide to social science data preparation and archiving.* (n.d.). ICPSR; University of Michigan. Retrieved January 29, 2024, from <https://www.icpsr.umich.edu/files/deposit/dataprep.pdf>

*Help and user guide.* (n.d.). ICPSR; University of Michigan. Retrieved January 30, 2024, from <https://www.icpsr.umich.edu/web/about/cms/5078>

*ICPSR*. (n.d.). ICPSR; University of Michigan. Retrieved January 28, 2024, from <https://www.icpsr.umich.edu/web/pages/index.html>

*List of member institutions and subscribers*. (n.d.). ICPSR; University of Michigan. Retrieved January 28, 2024, from <https://www.icpsr.umich.edu/web/membership/administration/institutions>

*Metadata*. (n.d.). ICP; University of Michigan . Retrieved January 31, 2024, from <https://www.icpsr.umich.edu/web/pages/datamanagement/lifecycle/metadata.html>

Pragadeesh. (2021, June 21). 12 types of survey questions to collect all types of data. SurveySparrow. <https://surveysparrow.com/blog/types-of-survey-questions/#id1>

Quirin Manwiller, K., Peter , S., Anderson, A., & Crozier , H. (2023). Dublin Core Indexing Metadata for Hidden Barriers: The Experience of Academic Librarians and Archivists with Invisible Illnesses and/or Disabilities. College & Research Libraries, 84(5). <https://crl.acrl.org/index.php/crl/rt/metadata/26024/33938>

Quirin Manwiller, K., Peter , S., Anderson, A., & Crozier , H. (2023). Hidden Barriers: The Experience of Academic Librarians and Archivists with Invisible Illnesses and/or Disabilities. College & Research Libraries, 84(5). <https://doi.org/10.5860/crl.84.5.645>

Quirin Manwiller, K., Peter , S., Anderson, A., & Crozier , H. (2021, December 13). Hidden barriers: The experience of academic librarians and archivists with invisible illnesses and/or disabilities dataset. FigShare; West Chester University. <https://doi.org/10.15786/17161328.v1>

*Start sharing data.* (2018). ICPSR; University of Michigan. <https://www.icpsr.umich.edu/web/pages/deposit/index.html>

Text file format - what is A .TXT and how to open it. (n.d.). Adobe. Retrieved January 26, 2024, from <https://www.adobe.com/uk/acrobat/resources/document-files/text-files/txt.html#open>

*What are the differences between ICPSR’s deposit options?* (n.d.). ICPSR; University of Michigan. Retrieved January 28, 2024, from <https://www.icpsr.umich.edu/web/ICPSR/cms/3940>