

## Project Abstract

Data about public libraries is becoming more prevalent but is still slow to be published. As publicly funded institutions, public libraries are also facing a constant struggle to obtain their portion of annual budget allocations. The data presented here are curated to address this problem by working with available data to highlight specific ways in which libraries can benefit their communities and to justify their budgets. The intended audience for the data is policymakers responsible for making budgetary decisions for public libraries. The hope is that the data can be used to showcase the many attributes of public libraries and convince policymakers to increase library budgets to better serve their communities.

This dataset was created to offer policymakers the ability to quickly evaluate the most impactful services offered by rural public libraries. This was achieved by compiling data collected in 2019 from the Library Search and Compare tool created and maintained by the Institute of Museum and Library Services. This data was curated by choosing to include data for fifteen libraries defined as rural by the IMLS. There are a total of seventeen libraries the IMLS defines as rural, but two were removed during the data cleaning process as they are part of larger library districts and have significant differences in revenue and service population.

A heterogenous dataset was curated by selecting thirteen variables that highlight the variety of services rural libraries offer, and to show the potential uses for budgetary increases. These variables were also chosen to give the most relevant information to policymakers to aid in budgetary decisions. The decision was made to store all data and protocol documentation in the open access GitHub repository named *Washington\_Rural\_Libraries*.

The main challenge in curating this data was working with the IMLS definition of “rural” which is in accordance with the U.S. census definition of “any population, housing, or territory not in an urban area” (*U.S. Census*). This definition makes it challenging to understand the distinct differences of each rural location, as it simply groups all non-urban areas together. This definition also limits the ability of comparing data across resources as other sources define rural differently. For example, the Library Research Service defines rural as any service area population under 25,000 (LRS). This difference in definition makes it difficult to offer any standard comparisons between data sources. A future challenge for and potential use for this data is incorporating it with other definitions of rural to better understand the challenges these libraries face and how meaningfully curated data could help to display the impact of rural libraries.

## Documentation

### Metadata

Schema: Dublin Core

Attribute	Value
Spatial Coverage	Washington
Temporal Coverage	2019

Creator	Tara Lanphar
Date Created	2022-03-04
Description	The data represent a range of statistics for rural public libraries in Washington and were curated to detail the impact of these libraries for policymakers responsible for budgetary decisions.
Format	CSV
Identifier	<a href="https://github.com/tleigh206/Washington_Rural_Libraries">https://github.com/tleigh206/Washington_Rural_Libraries</a>
Language	en
Publisher	Tara Lanphar
Rights	This is an open access dataset available for public use without restrictions.
Source	<a href="https://www.imls.gov/search-compare/">https://www.imls.gov/search-compare/</a>
Subject	Rural Libraries
Title	Washington Rural Libraries
Type	Dataset
Audience	"Policymakers," "data curators"

### ***Readme File (WRL\_2022\_Lanphar\_Final\_Readme.txt)***

This readme.txt file was generated on 2022-03-04 by Tara Lanphar

#### ----- GENERAL INFORMATION -----

1. Title of Dataset: Washington Rural Libraries

2. Author Information:

Name: Tara Lanphar  
Institution: University of Washington  
Email: [taral2@uw.edu](mailto:taral2@uw.edu)

3. Date of data collection:

Data were collected from the Institute of Museum and Library Services (IMLS) Search and Compare tool on 2022-03-01.

The data collected from IMLS reflect library survey data collected for the year 2019.

4. Geographic location of data collection:

The clean data represent 15 rural libraries throughout Washington including the cities of: Cathlamet, Davenport, Harrington, Kittitas, Lopez Island, Odessa, Eastsound, Pomeroy, Reardan, Ritzville, Friday Harbor,

Sprague, Concrete, Waitsburg, and Wilbur.

5. Information about funding sources that supported the collection of the data:

IMLS is a federal agency funded through the annual federal appropriations process.

The curation of data for the Washington Rural Libraries 2022 dataset was not funded.

-----  
DATA AND FILE OVERVIEW  
-----

1. File Naming convention:

WRL\_YYYY\_Name\_FileState\_FileType

a. WRL refers to the title of the dataset (Washington Rural Libraries).

b. YYYY is the year the data were collected for the dataset.

c. Name is the last name of the file creator.

d. FileState may include any combination of these values as appropriate: raw, clean, draft, final.

c. FileType is any additional description about the type of information in the file, e.g., "protocol," "dataset."

2. File List:

A. WRL\_2022\_Lanphar\_Raw\_Data.CSV

Creation Date: 2022-03-01

This file contains the raw data downloaded from the IMLS Library Search & Compare tool.

B. WRL\_2022\_Lanphar\_Clean\_Data.CSV

Creation Date: 2022-03-04

This file contains the clean data that has been standardized to the specifications listed in the data dictionary.

C. WRL\_2022\_Lanphar\_Final\_Readme.txt

Creation Date: 2022-03-04

This file contains information pertaining to the entire Dataset. For the specific purposes of this assignment, the readme file is included within the protocol document.

D. WRL\_2022\_Lanphar\_Final\_Protocol.pdf

Creation Date: 2022-03-04

This file contains the data curation protocol for the Washington Rural Libraries dataset.

-----  
METHODOLOGICAL INFORMATION  
-----

1. There are no licenses or restrictions placed on this data.

2. Data is available at

[https://github.com/tleigh206/Washington\\_Rural\\_Libraries](https://github.com/tleigh206/Washington_Rural_Libraries)

-----  
METHODOLOGICAL INFORMATION  
-----

1. Methods of data collection:

Data were collected from the IMLS Search & Compare tool which can be accessed at <https://www.imls.gov/search-compare/>. This tool represents all data from the IMLS FY2019 Public Libraries Survey. This is a detailed survey collecting information for over 9,000 public library systems in the United States. The details of this collection process can be found at <https://www.imls.gov/research-evaluation/data-collection/public-libraries-survey>.

2. Methods for processing the data (data normalization):

The methodology for processing the data in this project is based on recommended data analysis protocols (LeTourneau). Data for this project were compiled from the IMLS Library Search & Compare tool. This tool provides multiple filter options and for this date the following filters apply:

State- Washington

## Locale- Rural

Raw data were then collected from the IMLS Library Search & Compare tool by downloading all available variables in a single CSV file. The data were analyzed to remove any extraneous variables that did not apply to the intended use of this dataset. The goal was to highlight library services and budgetary information including revenue and expenditures.

The selected variables are parsed and standardized for consistency with casing and alignment. Variables and attributes are organized into columns and rows, respectively. Variable labels should be in lowercase with spaces as underscores, while library and city names are in title case.

Data are cleaned to reflect the allowed values as stated in the data dictionary.

### ----- DATA-SPECIFIC INFORMATION -----

#### Data Dictionary:

Variable Label	Variable Name	Variable Type	Measurement Unit	Allowed Values	Description
name	Library Name	String	n/a	n/a	The name of the library
city	City	String	n/a	n/a	The city location of the library
service_area_population	Service Area Population	Integer	n/a	Any integer over 0	Number of people in the geographic area for which a public library has been established to offer services and from which (or on behalf of which) the library derives revenue, plus any areas served under contract for which the library is the primary service provider
total_staff	Total Staff	Decimal	n/a	Any number over 0	Total paid employees (full-time equivalent)

total_revenue	Total Revenue	Integer	Dollars	Any integer over 0	Annual Sum of Local Government Revenue, State Government Revenue, Federal Government Revenue, and Other Operating Revenue
total_expenditures	Total Expenditures	Integer	Dollars	Any integer over 0	Annual sum of all operating expenses for the library including staff and collection.
hours	Hours	Integer	Hours	Any integer over 0	Number of hours per year the library is open to the public
visits	Visits	Integer	n/a	Any integer over 0	Total number of persons entering the library for whatever purpose during the year
users	Users	Integer	n/a	Any integer over 0	Total number of people registered to use the library
total_circulation	Total Circulation	Integer	n/a	Any integer over 0	Total annual circulation of all library materials of all types, including renewals
total_programs	Total Programs	Integer	n/a	Any integer over 0	Number of planned events for the year which introduce the group attending to any of the broad range of library services or activities or which directly provides information to participants
program_audience	Total Program Attendance	Integer	n/a	Any integer over 0	Total number of participants that attended library programs for the year
computers	Internet Computers	Integer	n/a	Any integer over 0	Number of computers with internet access that may be used by patrons
computer_uses	Computers Uses	Integer	n/a	Any integer over 0	Number of total computer uses per year for all computers
wifi_sessions	Wi-Fi sessions	Integer	n/a	Any integer	Number of total wireless

					sessions per year for all computers with internet access
--	--	--	--	--	--

### ***Justification***

The Dublin Core metadata schema was chosen to display an appropriate amount of specificity to assist in the reuse of this data. Dublin Core provides a good balance of expressivity and tractability with those attributes necessary for effectively conveying information both for researchers wanting to reuse this data and for the intended audience of policymakers to better understand the information presented here. Dublin Core recommends fifteen general fields, and all were included except the “contributor” field as it did not have a value in this instance. The “audience” field was added due to the importance of identifying who these data are being curated for.

File naming conventions were decided in accordance with accepted best practices for this domain (*Stanford*). It was decided to use the year of file creation to enhance the findability of the documents relative to this specific assignment. The file\_state and file\_type work to provide version control throughout the process.

The readme file was completed based on the standards recommended by the Research Data Management Service Group (*Cornell*). The main areas of focus include a description of how the data are standardized, recommended file naming conventions, and the description of variables using a data dictionary. A data dictionary was chosen as the best method for displaying variable information as it allows for a thorough description of the variables. This will be important to policymakers to best understand the information while also relaying critical information to other data curators or future researchers.

### ***Reflection***

The curation of this data was straightforward in the sense that IMLS provides clear documentation and has created an intuitive tool with a wide range of relevant statistics around public libraries. The data were downloaded in an accessible format as a CSV file and there was little cleaning required as the data were mostly standardized. The difficult aspect of this data curation process was in achieving the desired purpose for this dataset. This problem arose from the lack of a clear definition of “rural” libraries. While the Search and Compare tool made it easy to limit the data to rural libraries in Washington, this data was restricted by the unsatisfactory IMLS definition of rural as a general “non-urban” category. The only way to compensate for this restricted definition was to limit the seventeen rural libraries to fifteen based on the two libraries being outliers in terms of revenue and service area.

While this limitation reduces the scope of this project to the IMLS-defined rural libraries, it presents an opportunity for future data curation projects that could take a deeper look into how rural libraries should be defined, and what variable are most meaningful in determining how libraries are able to provide services. This reflects the initial problem defining this project in that there is an overall lack of current data around libraries in in general. For example, it would be informative to compare libraries within the “rural” designation by only total revenue or service area population. This would further normalize the data and provide meaningful

insight into how similar libraries are functioning and what services different revenues allow for libraries to provide.

Overall, this was a challenging yet rewarding project. It highlighted the multi-faceted nature of data management and how difficult it can be to ensure the proper information is conveyed to assist in the future use of the data, and to help the audience utilize this data for the purpose in which it was intended. Every decision had multiple potential solutions, from file naming conventions and metadata schemas to documentation and repository choices, and each choice had to be weighed to determine the potential impact on the future use of the data.

## **Resources**



*Budget.* Institute of Museum and Library Services.

<https://www.ims.gov/about/mission/legislation-budget/ims-budget>

Dublin Core Metadata Initiative. <https://www.dublincore.org/specifications/dublin-core/dces/>

*Library Search and Compare.* Institute of Museum and Library Services.

<https://www.ims.gov/search-compare/>

*Public Libraries Survey.* Institute of Museum and Library Services.

<https://www.ims.gov/research-evaluation/data-collection/public-libraries-survey>

*readme.* Cornell University. <https://data.research.cornell.edu/content/readme>

*Research Guides.* LeTourneau University. <https://lib-guides.letu.edu/quantresearch>

*Rural America.* United States Census Bureau.

<https://mtgis->

<portal.geo.census.gov/arcgis/apps/MapSeries/index.html?appid=49cd4bc9c8eb444ab5>

<1218c1d5001ef6#:~:text=The%20Census%20Bureau%20defines%20rural,tied%20to%20>

<the%20urban%20definition>

*Rural and Small Public Libraries.* Library Research Service.

<https://www.lrs.org/rural-and-small-public-libraries/>