# SPONSOR

Our Capstone project is sponsored by the Washington State Library, which, in partnership with the state government, maintains maintains data.wa.gov, the official Washington state government open data portal.

The project sponsor is Kathleen Sullivan, Washington State Library Open Data Librarian.

# OUR TEAM

Our team is made up of four MLIS students at the University of Washington Information School.



## MARIA ARTEAGA CUEVAS

Outreach Lead

Maria brings experience in data management and data curation to this project.



## SAM MINTZER

Project Manager

Sam brings experience in project management, comprehensive literature review design, and data curation to this project.



## REBECCA STEWART

Literature Review Lead

Rebecca brings experience in design thinking, collection development, and data curation to this project.



## CALEB SAUSSY

Deliverable Lead

Caleb brings experience in collection development, literature review, and policy design to this project.

# INFORMATION NEED

- The Washington State Library (WSL), which is part of the Office of the Secretary of State, maintains an open data portal as one of its library collections.
- Currently, the portal is populated with data provided voluntarily by government agencies, but data selection and publication are not guided by documented criteria.
- The WSL team has identified a need for firmer open data selection practices.
- Through a comprehensive, systematic review of peer programs and relevant literature, we will identify high-value sources that WSL can draw on to outline goals and practices for open data selection.
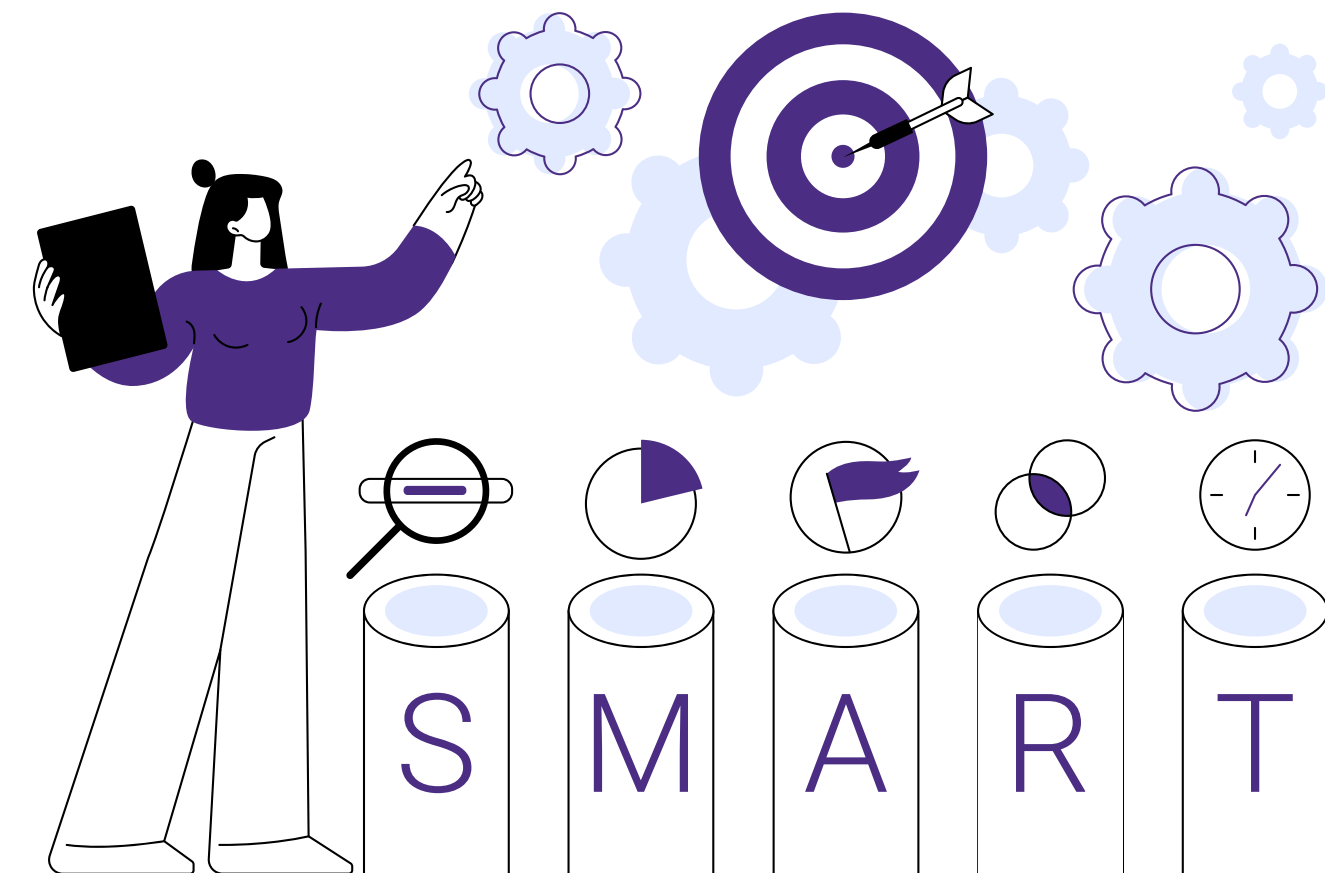
# INITIAL OBJECTIVES

**1** Conduct a systematic review of open data organizations and their open data selection practices.

**2** Conduct a systematic review of academic literature that address open data selection practices.

**3** Contact open data organizations and conduct informational interviews that can help reveal data practitioners tacit selection strategies.

SMART

# EMERGING THEMES

Through our systematic review and interviews, we identified a set of themes related to open data selection practices in the United States. State and municipal government organizations take different approaches regarding the following:

**3** **Transforming Data into Information**
Data is high-value if it can be used to answer questions for stakeholders. This theme refers to practices used to deliver data to users in information-rich formats, like data stories and visualizations.

**1** **Defining Data Roles**
Data roles describe all the people responsible for data throughout its lifespan - from creation to curation to publication.

**4** **Establishing Criteria**
Practices and documentation used to identify and prioritize higher-value data for publication. Portals also refer to criteria when excluding certain data.

**2** **Performing Outreach**
Data is only useful if people use it! Outreach encapsulates the practices used to engage public and governmental data users as well as data publishers.

**5** **Determining Selectivity**
Programs have different ideas about the scope of data that they want to collect. This theme refers to practices for modulating that scope.

# RESULTS



**Data selection is relational:** Data comes to open data programs from data-producing government agencies.
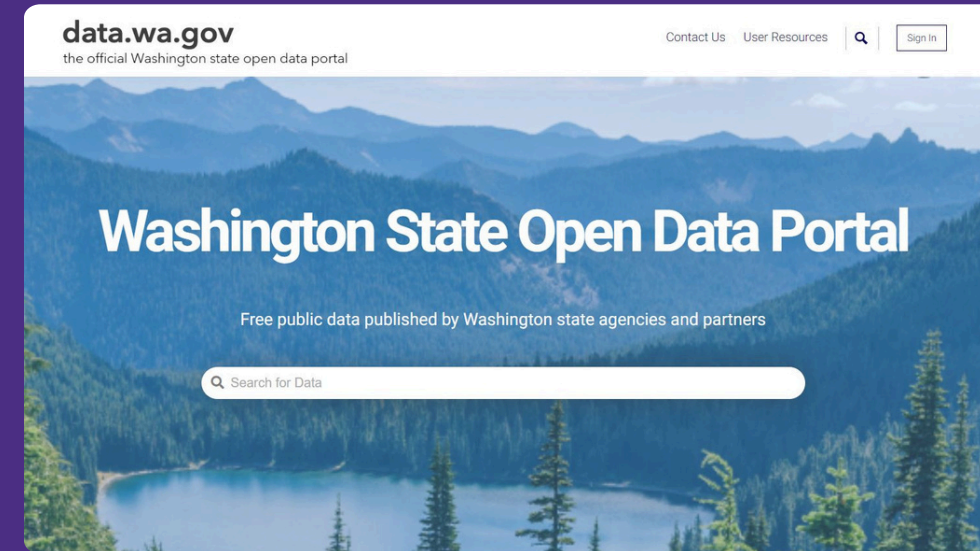
- Ultimately selection is a matter of managing relationships with those agencies, whether through outreach, education, or evangelism.

**Data selection is the wild west:** Opening data can be messy; there's a lot of it!

- Best practices are often a lower priority than simply getting all the data published.

**Data is contextual:** There's so much variance in data that the best practice for a dataset is often particular to that dataset.

- The most effective practice for increasing data value across the board is to have open data teams educate data creators to make them more effective data sharers.



**Memphis, TN**
*Summary:* Data champions, representing different data-producing agencies, work in conjunction with the Mayor to select data to publish. A data governance committee is responsible for reviewing overall policy direction. The mayor is given final say on what data is published—this is a top-down governance structure. Regarding public participation, two public representatives serve on the data governance committee, and the portal does allow for datasets to be suggested; there is some emphasis on creating opportunities for public participation in the executive statement for the program.

*Themes:* Data management roles  Outreach

*Reviewer Note:* We see an agency representative (aka champion) model. Having a single figure serve as primary decision maker, in this case the mayor, is noteworthy. Soliciting public representatives for the governance committee (outreach) is also a distinct feature.

*Sources:* Executive Order (last updated 2022)

**Minneapolis, MN**
*Summary:* Minneapolis sets 3 requirements for dataset prioritization: "For purposes of prioritizing data sets for inclusion on the Open Data Portal, Departments should consider the following: (1) whether the data is frequently requested; (2) helps increase the City's accountability, efficiency, responsiveness or delivery of services; (3) the data set improves public knowledge of the City and its operations, or furthers the goals of the City." One data coordinator is assigned to each data producing agency to help open up data, and the chief information officer oversees policy implementation. There is no clear avenue for user input.

*Themes:* Data management roles  Criteria

**Themes**
5 Government/Broader policy direction,
3 Criteria
3 Selectivity
2 Data Management Roles
2 Data to Information
2 Outreach
1 Public Interest
1 Collaboration
1 Automation

# NEXT STEPS

- Deliver a summary of our findings to both our sponsor and organizations that participated in our interviews.
    - This summary will include a breakdown of notable data selection trends and unique practices.

- Make recommendations for areas that Washington State Library may want to conduct more research on, including AI/ Automation practices in open data environments.

- Provide our sponsor with documentation that describes the scope of our systematic review. This will allow staff or future capstone students to iterate, not duplicate!