Selection Practices in Open Government Data Publishing

Maria Arteaga Cuevas
University of Washington
arteamar@uw.edu

Sam Mintzer
University of Washington
smintz2@uw.edu

Caleb Saussy
University of Washington
csaussy@uw.edu

Rebecca Stewart
University of Washington
rlstew@uw.edu

Abstract

This investigative report, sponsored by the open government data program at the Washington State Library, delves into the practices of prioritizing, evaluating, and determining publication priority at various open data government portals. Drawing on selection policies from peer organizations, recent academic literature, and interviews with open data representatives, the report outlines methodologies, major themes, and trends in open data selection policies. It highlights key practices such as scoping and selectivity, data management roles, public interest considerations, agency relationships, policy direction, criteria establishment, and emerging trends like AI integration and cross-portal integration. The report underscores the relational nature of data selection, the contextual nature of best practices, and the ongoing evolution of open data publishing as a field.

1. Introduction

This report is the product of an investigation, sponsored by the open government data program at the Washington State Library (WSL), into the practices used to prioritize, evaluate, and determine publication priority at other open data government portals (described here as selection practices). Our research explores the current landscape of the highly emergent field of open data selection policies, and makes provisional recommendations based on our findings.

Our research takes into account three sources of information about open data selection: selection policies published by peer open data organizations, recent academic literature that addresses open data selection, and interviews with representatives at open data organizations. Through this three-stage, comprehensive review process, we have identified a broad range of current practices and concerns across the open data field.

This document outlines our project methodology, the major themes in open data selection policies that we identified, and some analysis of trends, outliers, and other findings that may inform future policy direction at the WSL.

2. Description

Open Government Data, at this time, remains an emergent field in which best practices and target outcomes do not yet conform to a clearly established consensus. In the absence of such a consensus, publication practices are often developed based on need and remain internal to the programs where they originate; programs may not even consider them to be publication practices as such. The framing of our survey is to use the tradition of library collection development—a more mature discipline with better-articulated concepts and language for making information accessible to the public—to survey, analyze, and catalog the practices through which publishing programs currently practice selection.

In this report, we use the language of traditional library collection development to describe the nature of open government data publishing. **Selection**, in collection development terms, describes the full range of practices responsible for determining how items enter the collection, as well as how a target collection is determined. In the context of a data portal, selection practices are often restricted to what libraries might call **gift management**; that is, published data ultimately comes from **donors**, or data-producing agencies, and the selector's role is to determine the value of incoming **gifts** (data) and manage relationships with donors. Unlike in the traditional collection development context, data selectors rarely reject gifts; instead, successful data publishing frequently centers on **donor relations**, or the outreach, education, and relationship maintenance necessary to pre-emptively ensure that data is suitable for publication. The vocabulary used by institutions responsible for creating and opening data is inconsistent across different contexts; this report uses the term **agency** to refer to organizations that act as

data donors and **publisher** to refer to organizations responsible for collecting, curating, and disseminating that data.

Although there is overlap between traditional library collection development and open government data selection, we make the crucial distinction that data publishing is an ongoing, constant, "live" process whereby data is continually pushed to a portal, unlike traditional library collection development, in which items are published once and rarely undergo further change. Selection for open data publication, therefore, also entails the prioritization of data for ongoing publication. We use the term **high-priority data** to describe data determined to be of especially high value to the public interest or otherwise worth publishing over other data.

3. Methods

To survey the field of open data publishing, we employed three avenues of research: a review of recent academic literature, a review of data policies at peer programs, and interviews with those programs. Full notes on our inclusion/exclusion criteria, our lists of peer organizations, and our interview protocols have been delivered to the WSL.

Open government data is a broad and complex field, and the primary challenge of our process was determining which peer programs to investigate. Given the sheer volume of government managed open data programs in the U.S. we began by creating an inclusion/exclusion criteria that could help us limit the sample of sources we would be reviewing (detailed in our inclusion/exclusion criteria document). We identified a list of open data programs which are frequently updated and feature mature policy documentation relevant to what we've described as open data selection, which we used as context to scope our three categories of research

3.1 Academic Literature Review

Our academic literature review process was surprisingly difficult. Our search criteria were designed to be overwhelmingly broad: we reviewed information science, data, and policy research databases for a number of keywords we identified as synonymous selection practices (ie. data workflows, curator responsibilities, or quality standards; full details are available in our documentation) related to open data, as apparent in the article abstract. We specifically excluded work older than five years, as open data is a rapidly evolving field, as well as extra-American sources and those focused on processes deemed out of scope such as preservation or design. Although we did not expect the academic literature to be deeply productive, given that the field of open data remains novel, practices tend to reside among practitioners, and the idea of selection is itself an unusual framing, we still found surprisingly little material that met these broad criteria. We assume that this scarcity reflects both a division between academia and active practice in the field as well as a broad decline in academic interest, possibly driven by the demise of the Sunlight Foundation and other major institutional supporters of research on the topic. Ultimately, we reviewed 12 articles of broad relevance, which we employed as context and background for our other research.

3.2 Peer Organization Policy Review

Our review of peer organizations' policies included all possible documentation related to an organization's data program, including memoranda, policies, articles, handbooks, and accessible internal documentation. In short, reviews were based on a synthesis of materials, not any one policy document.

An initial sample of peer organizations was created based on two sources: A) Organizations that Sunlight Foundation had worked with in the past and archived policies for; and B) State organizations. The rationale for including A is that any state, city, or county that has worked with Sunlight Foundation in the past may have a more mature, robust data program today; checking in on these programs could yield insights. The rationale for including B is that WSL will have a strong interest in state data program policies. From here, peer organizations were included or excluded based on our inclusion/exclusion criteria. In total, 53 peer organization programs were deemed relevant.

Of these 53 organizations, 18 were reviewed and summarized. Summaries were constructed by looking at how organizations dealt with:

- data evaluation criteria;
- tools and methods used to select data;
- gift management;
- and end-user participation.

Summaries were then coded and reviewers offered initial thoughts on the strengths, weaknesses, and unique attributes of each program.

Time constraints limited our capacity to review all 53 organizations in our relevant sample. A more rigorous inclusion/exclusion criteria may have benefitted us

here. The challenge of synthesizing information about data selection from many sources within one organization was also underestimated. Future work could begin with examining some of the organizations that were not captured by our review.

3.3 Peer Organization Interviews

In order to gather a more complete idea of the open data domain and active practitioners' methods of working with high priority data, we conducted a series of informational interviews. These interviews aimed to survey the state of government open data portals and determine the tacit practices or tools being used to choose data to prioritize for publishing. These interviews were especially important to our research, as they provided us with the most direct and meaningful access to the practices actually being used in the field today. Our interview candidate list included 21 state and local government publishers and their representatives, which were each selected for a specific reason. Reasons included the presence of mature open data programs, extensive documentation, standards or rubrics for data selection, or a recommendation from our project supervisor. We were ultimately able to secure interviews with one third of our candidate list.

Although these information interviews were informal, we created an interview protocol to standardize our questions and the background and context we shared with interviewees, attached in our documentation. This protocol was designed to ensure that all participants had an equal understanding of the goals and intentions of our project.

The interviews were held virtually over the span of three weeks and were scheduled to last a minimum of 30 minutes. Two team members conducted each interview, one responsible for taking notes and one responsible for leading questions. We recorded interviews with the explicit consent of our interviewees, and used those recordings for transcription in order to supplement our notes; afterwards, recordings were deleted.

State and local data portals operate very differently, and this occasionally posed a problem during the interview process, as some of the questions were not immediately relevant to all interviewees. Although awkward, this allowed us to gain a better understanding of each agency's priorities and involvement with data selection. Time constraints also worked against us in this area as we were working with the 10 week quarter at the UW and a small and complex window of availability from all group members. Ultimately we gathered that for future efforts, expanding our interview windows to a minimum of 45 or 60 minutes would allow for a more meaningful interview experience.

3.4 Methods - Conclusion

After completing the three research processes outlined above, we employed qualitative coding across our literature review and interview notes in order to distill the wide range of practices we'd identified into categories and make our best effort at identifying best practices. Our full codebook can be found in the attached documentation.

4. Trends & Recommendations

Through the research process described above we identified a number of trends into which we were able to group the various practices used by publishers to select data. Although these categories do have a certain amount of inevitable overlap, and some of the practices we describe below might easily fit into two or three separate categories, we've established eight thematic groups of practices that we consider to be of interest to the open government data selector.

4.1 Scoping & Selectivity

The essential question, when making selections, is that of the target collection's scope - what and how much data ought to be published. Each program conceives of its scope differently, and programs use a variety of different tools to modulate that scope. The most obvious of these is policy - scope is quite clear when it is defined by law (as it is in Texas) or by policy (for example, the city of Seattle's new open-by-default data policy). However, scope is also tacitly enforced through a variety of other policies and practices - metadata and technical requirements, the development and implementation of strategic goals, and governance structures all work to define a program's scope and target collection.

These practices are most obvious when looking at programs with unusual scoping objectives, which apply correspondingly selective thinking in their selection decisions. The state of California's open data portal, for example, thinks of itself primarily as a complete co-location for all the state's data (distinguishing itself from other, more specialized data portals in the state) and therefore prioritizes the acquisition of data above all else - even when that data has technical

compatibility issues or incomplete metadata. On the other end of the spectrum, the city of Syracuse, New York, publishes data only if signed off on by the city's data governance committees. Neither approach is better than the other; what's allowed both those programs to succeed is a conscious and explicit articulation of the scope of the publisher's target collection which can then be pursued through the development of policy and governance.

Highlighted Example:

- For Baltimore, MD, the onus is on data stewards to evaluate data on an Impact vs. Effort scale to decide what's worth publishing. Even if a dataset might be high-value, if it isn't high-quality and would require great effort to curate and maintain, stewards are encouraged to consider moving on to something else (Brandon et al., 2022). Baltimore goes further in this document and actually breaks down what the "costs of bad data" are. Time and money are considerations when determining what data should be opened by preference.

4.2 Data Management Roles

Organizations assign roles and duties to individuals in the data publishing pipeline. Across the range of publishers we surveyed, there are variations in terms of how those roles are defined. Specific titles include Open Data Champions, Data Stewards, and Agency Data Officers. Though specifics in roles may change by state and agency, these often describe a designated individual from each agency that acts as a liaison and the point person for open data in that part of the organization (N. Diaz Amigo & J. Scharf, personal communication, April 17, 2024), (M. Schmidt, personal communication, April 16, 2024), (P. Zaldonis, personal communication, April 29, 2024).

Maintaining clear data management roles is an important tool in data selection. Relationships should be maintained with Open Data Champions, Data Stewards, and /or Agency Data Officers as they will be most aware of the data that their department holds and can communicate that information back to the open data librarian. Data management roles are also important when it comes to interagency relationships. Establishing and maintaining open data management roles at partner agencies can be a key step in setting publishing expectations, like prioritization or metadata needs, that agencies often find difficult to enforce.

Highlighted Examples:

- In an interview with leadership at Syracuse Open Data representatives discussed that they have a data governance group made up of one data steward from every city department, data focused staff, and their API team. Additionally they have a data steering committee with higher-level executives. In order to share data, the data governance group votes on the matter, and if a decision can't be reached the vote goes to the data steering committee. For this open data portal, datasets have been coming directly from the data stewards as they are typically responsible for maintaining a data inventory for their department. Syracuse notes that the committees are critical when determining what data is of high value and why it should be published (N. Diaz Amigo & J. Scharf, personal communication, April 17, 2024).

- The state of California's Open Data Publisher's Handbook recommends establishing several different roles, each with its own level of involvement, to ensure a comprehensively established chain of custodianship throughout data's life cycle at an agency (California Open Data, 2024).

4.3 Public Interest

The public (that is, non-government actors) are a crucial user audience for the publishing/opening of data. Their interests, in the absence of better tools or outreach, are generally made known to publishers primarily through FOIA or FOIL disclosure requests. These requests are sometimes communicated through built-in features within portal websites, such as "request a dataset" forms. Other times a more formal method is used where the public communicates interest in opening data through Freedom of Information Act requests.

Highlighted Examples:

- In the case of Glendale, AZ, public data requests are used to generate more nuanced models of public interest. Glendale, in partnership with the Sunlight Foundation, based on information requests, generated user personas for individuals interested in real estate data. These personas were used to determine the types of datasets that real estate stakeholders are interested in. Glendale and Sunlight then returned to stakeholders to continue conducting interviews about what data are most impactful. Here a combination of value-sensitive design factors (stakeholder interviews, persona development) are used to get a better sense for what public interest in data is. (Sunlight Foundation, n.d.)
- New York City maintains a public information desk where citizens can make requests, which is specifically differentiated from the city's review of FOIA requests in that it's oriented towards

ongoing, rather than one-time, publication. (Z. Feder, personal communication, May 3, 2024)

- State data portals are meant for everyone to be able to access, and public interest should therefore play a role in what data is selected to be on the portal. Anticipating the interests based on previous PRA/FOIL requests can lessen the amount of these requests coming in. For example in our interviews the state of Texas mentioned they review FOIL requests to gauge public interests in data and have seen a reduction of FOIL requests when they open up their data based on that information. (N. Cooke et al., personal communication, April 23, 2024)

4.4 Agency Relationships

This category describes practices on the part of a publisher to solicit data submissions and establish a publishing relationship with agencies. This category also includes interagency collaboration, which can be a powerful tool to increase the standardization and understanding of data. Publishers have a wide variety of tools to manage such relationships, including the establishment of governance committees and the creation of reference material and training for agency partners. Since much of what constitutes open data selection happens at the agency, before it reaches the publisher, the base of understanding built by these relationships are one of the most significant tools at a selector's disposal. As an example, New York City has credited their success to a strong "data culture" fueled by extensive experience, collaboration, and use of data across agencies at multiple levels (city, state, etc.).

Highlighted Examples:

- The California Open Data Publisher's Handbook provides a comprehensive reference on the data publishing process for partners, including setting expectations for metadata and data cleaning (California Open Data, 2024), (S. Hayashi et al., personal communication, April 23, 2024).
- The New York State Open Data Handbook provides a particularly thorough description of a data governance committee (under the vocabulary of "data working group".) Data governance committees are a fairly common governance model, especially on the East Coast and in smaller jurisdictions in which each agency's designated representative (data coordinator, in the New York State text) sits on a committee that makes decisions about data publication. Models like this one ensure stakeholder engagement in data publication, although they do rely on a baseline level of interest from agencies (data.ny.gov, n.d.), (Z. Feder, personal communication, May 3, 2024).
- Regularly scheduled training or outreach sessions can help maintain strong relationships with data stewards for data-producing agencies, thereby standardizing different publishing and inventorying processes. Through our interview, we learned that this is something that NYC has experimented with. Other data organizations, like New Jersey, also attempt to "game-ify" outreach to municipalities, encouraging them to adhere to data standards through a point system. All of these practices help strengthen interagency relationships (Sustainable Jersey, n.d.)

4.5 Policy Direction

Decisions on data selectivity and portal goals can sometimes be driven by specific policy direction. This category includes the practices of deferring to special legislative interests, strategic plans, legal requirements, institutional missions and goals, or emergency situations (e.g. wildfires, Covid-19), among others.

Highlighted Examples:

- Tempe, AZ, has created a "data feedback loop" that is responsive to broader government directives. Wastewater analytics are being used to better understand opioid use and the effectiveness of government interventions that have been put in place to address the crisis. Reducing opioid abuse is a key executive goal for the city, and by making community-health data a priority, open data coordinators can see the effectiveness of policy while highlighting where policies haven't been as effective. Data selection is driven by a top-down desire to tackle substantial issues with data-drive solutions. (Tempe, n.d.)
- NYC's data selection is in part guided by law. Data-producing agencies are expected to maintain and make available data that are considered public in line with NYC's Open Data Law. Compliance with NYC regulations is important, and preparing annual compliance plans requires close cooperation between data-producing agencies and the Open Data Team (City of New York, NYC Open, n.d.), (Z. Feder, personal communication, May 3, 2024).

4.6 Data to Information

One definition of high-priority data might state, reductively, that data is of high value when it can be used to answer questions of interest. Information-rich formats, like data stories and visualizations help bring data to users. Some agencies have been able to implement more advanced practices for prioritizing high-information datasets; the state of California, for example, runs annual hackathons inviting the public to build applications with their water data, and has used

feedback from those competitions to develop their understanding of the public's information needs. Prioritizing use cases like the development of data stories can help a selector ensure that the data in their portal serves the public interest. Selection here also means rejecting data. The city of Syracuse, for example, described receiving and implementing a data request for road temperatures from a local college, which the publisher ultimately had to scrap because said data was not in use by the government at any level - designating it as low-value information.

Highlighted Examples:

- Sometimes "data to information" practices look like data curation. That is to say, an organization will collocate data on different topics into one hub. This hub can be used to address user needs. For example, Baton Rouge, LO has an Open Neighborhood BR that brings together data on police reports, fire reports, traffic incidents, 311 requests, and building permits into one searchable map. This is an informative format that is hospitable to users who may be interested in interactivity but may not be interested in downloading and using data in other formats.
- The state of Connecticut has racial equity explicitly labeled as a strategic goal in its State Data Plan (Office of Policy and Management, 2022). Explicitly prioritizing this class of user questions allows the portal to better support analytic use cases, including its own data stories (such as that on areas disproportionately affected by cannabis regulation) and those external to the publisher (such as the Connecticut Data Cooperative's Zoning Atlas).

4.7 Criteria

This category describes how organizations use internal documentation to evaluate and prioritize data submissions. Criteria are also used to assess data quality considerations, such as metadata quality, as well as legal restrictions on data. Criteria can be informal (e.g. stated principles or strategic vision for data) or formal (e.g. scorecard systems).

Highlighted Examples:

- New Orleans' Open Data Portal scores all incoming datasets along well-defined criteria (see "How To Score Data" in Training - DataDriven - City of New Orleans) (DataDriven - City of New Orleans, n.d.). Many of the guidelines on this page are adapted from San Francisco's open data resources. Scorecards are filled out by data stewards for data-producing agencies, and stewards are encouraged to work with Data Coordinators if they run into difficulties. Data are scored along 3 metrics: Data Quality, Open Data (i.e. privacy), and Data Costs). Data stewards are then encouraged to use scorecards as a way to select data to publish and to set goals (e.g. the department will publish X guidelines in Y months). Here, scorecards help determine publication priorities.
- NYC frames selection criteria as a 6-point priority list. Though this isn't as strict a form of criteria as a scorecard, it looks at data through the lens of strategic goals. For example, the priority list asks if incoming data will create equitable access to public resources, respond to a need, or improve public knowledge of an agency. These questions are effective at framing selection as something that should always contribute directly to the goals of NYC's data program and its requirement to serve the public. Criteria, in this scenario, are used to rearticulate strategic priorities into a set of actionable selection metrics (NYC OTI & NYC OpenData, 2023).

4.8 Emergent Practices

Open government data, as we have noted more than once already, is an emergent field. Some practices are too new or underdeveloped to evaluate, but we will outline them here. First, we have encountered some discussion of and excitement regarding the use of artificial intelligence to parse and interpret data for users. The sole implemented example we're aware of, at time of writing, is the District of Columbia's "Compass" AI. It's unclear what role AI might play in the future of data publishing.

A second emergent practice is increasing or improving the use of automation to update or modify data with little human involvement; this is something a few different portals have described as part of their longer-term goals for the organization.

A third emergent trend is increasing interest in cross-portal integration, especially between different levels of government. For instance, the open data portal for the state of California may have reached a "saturation" point, since it now collects the majority of obtainable and publishable government data; as a next step, it has speculated on integrating county and city portals into the state portal.

It is unclear what kind of recommendations to make about these trends to a selector other than to monitor their development. We might, very speculatively, suggest that something like AI or automation could one day have ramifications for publishers' metadata management.

5. Discussion

To summarize and conclude our report, we offer the following three overarching synthetic insights drawn from the above analysis of practices in the open data field:

5.1 The selection of data is an inherently relationship-driven enterprise.

Data comes to a publisher from an agency, and selection is therefore influenced most heavily by what that agency provides. How a selector manages that relationship, then, is a critical piece of selection practice. To return to the language of library collection development, publishing open data is gift management, and the open data selector works primarily through donor relations. The most consistent and pertinent concern we've noted across our interviews with peer programs has been the establishment and maintenance of those relationships. Most of what we've cataloged in this report is ultimately an artifact of those relationships—things like governance structures, assigned liaison roles, policies, and agency outreach programs are all expressions of the open government data publishing field's reliance on establishing and maintaining agency/publisher relationships.

5.2 Best practices are contextual, and standards remain an ongoing challenge.

Data is messy, and what works in one context may not work in another. Defining and managing the policies, governance, and outreach that go into publishing a mandated annual report of agency activity, for example, poses a selection challenge altogether different from managing a constantly-updated flow of GIS data. Likely as a result of this, we only rarely came across standardized tools like scoring matrices for evaluating data selection. Some of the programs we investigated have mentioned moving away from those approaches due to the difficulty of enforcing them consistently and in a way that ensures the data is useful.

Something that can be standardized is the set of expectations provided to the partners of an open data publisher. New York City's program finds itself uniquely challenged with a huge variety of reporting agencies and legacy data models. To better navigate such complexity, the program has moved its guidance and documentation model away from standardizing technical requirements (like metadata) to instead focus on setting agencies' expectations about how the data should be made usable for the public; it currently requires only a data dictionary and a user guide.

5.3 Open data publishing remains a novel and unsettled field.

If there is one constant we have seen across our entire investigation, it is that publishing open data is messy and will continue to be. Though we have documented a number of practices that have proven effective for their user groups, organizations do not hold widespread consensus on best practices. Contributing to this is the fact that publishers still lack consensus in terms of their understanding of scope, priorities, governance frameworks, or other elements related to managing their target collections; it is difficult to articulate best practices for selection when these core concepts aren't homogeneously defined and understood across organizations. Of our themes, two—scoping and data to information—are areas where a selector's clear and explicit articulation of priorities can have the greatest clarifying impact on the chaos of our field. Publishers can most powerfully influence the relevance and accessibility of their collections by first defining and communicating the specific scope and purpose of their collection and then deriving the practices appropriate for achieving those goals.

6. References

6.1 Bibliography

California Open Data. (2024a, April 10). California Open Data Publisher's Handbook. https://docs.data.ca.gov/california-open-data-publishers-handbook/

California Open Data. (2024b, April 10). Review the Pre-Publishing Checklist | California Open Data Publisher's Handbook. https://docs.data.ca.gov/california-open-data-publishers-handbook/1.-review-the-pre-publishing-checklist

City of New York, NYC Open. (n.d.). NYC Open Data. Retrieved June 11, 2024, from http://nycod-wpengine.com/

DataDriven—City of New Orleans. (n.d.). Retrieved June 11, 2024, from https://datadriven.nola.gov/training/

data.ny.gov. (n.d.).

NYS Open Data Handbook | State of New York. Retrieved June 11, 2024, from https://data.ny.gov/dataset/NYS-Open-Data-Handbook/id8k-natf/about_data

NYC OTI, & NYC OpenData. (2023). Technical Standards Manual. https://opendata.cityofnewyork.us/wp-content/uploads/NYC_OpenData_TechnicalStandardsManual.pdf

Office of Policy and Management. (2022, February 28). State Data Plan 2023-24. https://portal.ct.gov/datapolicy/-

/media/datapolicy/connecticut-state-data-plan-2023---2024-final.pdf

Sunlight Foundation. (n.d.). Improving access to information on neighborhood development—Roadmap to Informed Communities. Retrieved June 11, 2024, from https://communities.sunlightfoundation.com/case-studies/glendale-development/

Sustainable Jersey. (n.d.). Retrieved June 11, 2024, from https://www.sustainablejersey.com/actions/?type=1336777436&tx_sjcert_action%5BactionObject%5D=537&tx_sjcert_action%5Baction%5D=getPDF&tx_sjcert_action%5Bcontroller%5D=Action&cHash=94657e52e56cabbc06d8bdc331ca9d65

Tempe, C. of. (n.d.-a). Tempe Wastewater BioIntel Program. Retrieved June 11, 2024, from https://wastewater.tempe.gov/

Tempe, C. of. (n.d.-b). Tempe Wastewater BioIntel Program. Retrieved June 11, 2024, from https://wastewater.tempe.gov/

6.2 Interviews

Cooke, N., Montgomery, A., & Rhode, B. (2024, April 23). Texas Open Data Portal [Zoom].

Diaz Amigo, N., & Scharf, J. (2024, April 17). Syracuse Open Data [Zoom].

Feder, Z. (2024, May 3). NYC open data [Zoom].
Hayashi, S., Wong, A., & Regan, K. (2024, April 23). California Open Data [Zoom].

Marvel, C., & Jerger, B. (2024, April 19). State of Alaska Open Data [Zoom].

Schmidt, M. (2024, April 16). City of Seattle Open Data Portal [Zoom].

Zaldonis, P. (2024, April 29). Connecticut Open Data [Zoom].