

INTERVIEW CONSOLIDATION: A THEMATIC DISTILLATION

Introduction

As our interviews crossed many government agencies and jurisdictions, we encountered several distinct perspectives from people concerned with public records and open data. Whether they were Public Disclosure Officers or open data officials, our interviewees helped us identify themes that cut across all agencies and jurisdictions.

Overall, organizations are hesitant to use public records as a source for proactive publication of open data. The tendency to be “risk averse” is a relevant tension that must be considered when undertaking an open data project. However, the participants did acknowledge that if such a proactive publication occurred, it could be possible that the frequency of public records requests would lessen and in turn, help reduce the cost of fulfilling such requests, whether that cost stems from FTE or potential risk.

We asked our interviewee's several questions about the possibility of participating in a regional Open Data Alliance. It is apparent that there is an overwhelming desire for a collaborative environment, an opportunity to better understand what each department or jurisdiction is doing related to open data and sharing best practices along the way that reinforces such a collaboration.

Themes

Public Records & Open Data

Skepticism about public engagement.

There is some skepticism about whether the public will engage with published datasets on an open data portal and if they are knowledgeable enough to work with the data in the form provided. Although this skepticism is valid, there is an inherent underestimation of the curiosity and skills among the public.

Data is not intrinsically useful by itself—one must engage with it to produce structured information, and later insight. There will always be some students or data scientists, hobbyist or community activists, that will find a way to use the datasets. For the parts of the public who might not know that such data is available or do not know how to work with the data directly, they can find aid with the help of their public library if they offer such services. However, this concern is equally an opportunity to educate the public and provide resources to support their initial curiosity.

Insights: Create partnerships between libraries in order to foster awareness about Open Data/Open Access materials produced by government.

INTERVIEW CONSOLIDATION: A THEMATIC DISTILLATION

Institutional embarrassment.

There is a hesitation to make particular datasets available because of the possibility that the data will not reflect well on the institution from which it came, and that “institutional embarrassment” will invite further public scrutiny for the institution. There is a constant drive to control the narrative of any situation, and this is heightened when dealing with projects that have a government-public interaction.

Potential embarrassment derived from the data, or even the intentional misrepresentation of the data is a possible risk associated with open data. An institution can preempt this by providing sufficient metadata to the datasets when publishing in the form of a data dictionary, explaining the source, collection methods, and so on. Understanding the context of the data collection can help mitigate the risks involved with potentially embarrassing insights.

Insights: Institutional embarrassment is a big barrier to open data. Focus on ways to alleviate these feelings by highlighting the good that open data can do.

Directing requesters to a dataset that is not often updated.

One apprehension about publishing frequently requested public records as datasets originate from a scenario where a Public Disclosure Officer instructs the public to the dataset in place of a formal request. If the data set is not updated regularly, there could be an issue where the requester did not receive all the documents available, and the organization would be liable legally. As such, it is crucial to make it clear how often the datasets are updated and by who and is integral to assuaging fears about such an issue.

Insights: If directing requesters to datasets in lieu of traditional public records request fulfillment, be sure that the data has a clear description of what it is and when it is updated.

Risks of inadvertently releasing personally identifiable information.

The mistake of publishing a dataset with personally identifiable information is of great concern. In particular, a case where one is automatically publishing a dataset, and there is not the opportunity to catch such an error until it is too late. Although the purely accidental release of personally identifiable information is possible, the occurrence of such an error is typically preventable by the department's privacy policies in combination with the public records officers who understand the exempt types of information.

One other possible way to prevent such an issue and to allow for the automatic publication of data comes from taking advantage of newly implemented data systems that replace legacy systems. For example, if a data collecting system in the office is receiving an overhaul or a rebuild, there can be a

INTERVIEW CONSOLIDATION: A THEMATIC DISTILLATION

moment of opportunity to restructure it in a way that will allow for a simple extraction of data that avoids all personal information.

Insights: When implementing or rebuilding systems that collect data, be sure to focus on ways that allow for easy data extraction and avoid tedious redaction.

Understanding the stewardship of legacy datasets.

In a government office, datasets are manifold. They are kept in many formats, housed in many locations, and are updated by employees, but it is not clear who owns them. For datasets to proactively publish or be published and continually updated, there must be an understanding of who owns the data and who is the person updating. While individual departments have a better understanding of their data stewards, others might not be on the same level. It can sometimes be common to find data for possible publication only to find that the person compiling the information has left or retired. There is still an opportunity to begin the stewardship anew and to attempt to discern the information as thoroughly as possible by compiling extensive metadata before publishing to an open data portal.

Insights: Tracking down data owners and data creators can be time consuming to do retroactively. Start from one point and continue to document so that future people working with or looking for the data will have some inkling of where to start.

Creating an uptick in requests.

There is a concern about the possible increase of public records requests when making other public records or datasets available as open data. The concern is rooted in the idea that such an uptick will only add to the increasing costs related to fulfilling public records requests, and that open data has not solved this issue but instead made it worse. Such a possible consequence of publishing open data is entirely new in our digital age because of the quick and accessible nature of the internet to the public and those concerned with government data.

It is difficult to determine whether an increase in public records requests is caused by the publication of data or with the implementation of a new open data portal. However, average increases in public records request often arise from incidences outside the department, such as cases when there is a police officer involved shooting and a subsequent increase in requests for their body-cam footage.

Insights: Increases in requests have occurred and will continue to occur regardless of open data publication. It is important to focus on the various datasets that can be published and their respective drops in requests, instead of the general overall public records request numbers.

INTERVIEW CONSOLIDATION: A THEMATIC DISTILLATION

Open Data Alliance

Creating a community of sharing around open data.

One hope for the Open Data Alliance, expressed by many, is the creation of a regional community that focuses on open data that isn't directly associated with the government but is made up of government workers. To really solidify the communal aspect of such an alliance, the voicing of opinions should be embraced by all members. Understanding that almost each person in this organization is familiar with the difficulties of publishing open data and utilizing that commonality as a basis for such a community will allow it flourish.

Insights: To be a successful organization, members must be open and accepting of all ideas.

Standardizing data to ease sharing.

Informal sharing and aggregation of data across departments and jurisdictions can be helpful in gaining insights, but the biggest issue is the standardization of this data. Unstructured, unclear data means more time spent preparing the data for analysis and takes up precious collaborative time if not addressed. Finding ways as a group to agree on an informal standardization would help ease sharing among the members of the Open Data Alliance.

Insights: Standardization takes time to create initially but is incredibly useful long-term.

A support system to learn/teach ways to work with data.

Taking advantage of the skills and education of the group would be extremely beneficial to ease the sharing of data and could result in people bringing good ideas back to their parent organization and affecting positive change.

The Open Data Alliance, for many interviewed, seemed an excellent way to support people who are working on a project related to open data or are interested in advocating for its use in their department. Sharing skills through peer teaching could create a support system for people new to the field. The goal of the alliance is to provide opportunities where one can be vulnerable enough to learn and strong enough to teach.

Insights: Utilize the knowledge and skills of all members and be sure to support members who are learning skills for the first time.