

INTERVIEW CONSOLIDATION: A THEMATIC DISTILLATION

Introduction

As the breadth of our interviews crossed many government agencies and jurisdictions, we encountered a variety of perspectives from various levels of management and interactions with public records and open data. From these interviews, we identified several themes that we felt intersected all agencies and jurisdictions. As public records requests represent a common thread among the interviewees, whether they be Public Disclosure Officers or concerned with open data, the themes presented are representative of their similar positions.

Overall, the themes formulated stem from the consistent hesitation of utilizing open data as an avenue for the proactive publication of public records. Not uncommon to government institutions, the tendency to be “risk averse” is an influential guiding factor when deciding whether to publish some public records as open data. 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.

As our project also focuses on the formalization of a regional Open Data Alliance, we asked our interviewee's several questions about the possibility of participating. From the responses, it is apparent that there is an overwhelming openness to the idea of a regional alliance. There is a desire for a truly collaborative environment, as well as an opportunity to gain an understanding of what each department or jurisdiction is doing related to open data.

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 underestimation of the curiosity and skills among the public. While it may not seem like some datasets are useful, there will always be some students or data scientists, hobbyist or not, that will find a way to utilize 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 possible aid with the help of their public library if they offer such services. However, such a concern can be an opportunity for an organization like the Open Data Alliance, to educate the public and provide resources to support their initial curiosity.

Institutional embarrassment.

There is a hesitation to make specific datasets available because of the possibility that the data will not reflect well on the institution from which it came. There is an additional concern that such embarrassment will invite further scrutinization of their institution.

While this is a valid concern, potential embarrassment derived from the data, or even the intentional misrepresentation of the data is a possible risk associated with open data. However, there can be an attempt at preempting such an issue by providing sufficient metadata to the datasets when publishing. Understanding the context of the data collection can help mitigate the risks involved with potentially embarrassing insights.

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.

The repercussions of such an error can be substantial and involve litigation. It is this concern that directly relates to the issues with understanding data stewardship of open datasets. 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.

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. By utilizing these rebuilding points, one could suggest that the development of a new system where data extraction can occur without having to redact personal information heavily.

Understanding the data stewardship of legacy datasets.

Depending on the department, there are datasets housed in many formats and locations that 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 thoroughly compiling metadata before publishing to an open data portal.

Creating an uptick in requests.

There is, understandably, a concern about the possible increase of public records requests when making other public records or datasets available as open data. The idea that the possible

attention or extra “eyes” on the department from which the dataset came, is one genuine possibility when engaging in the process of proactive publication.

Such a possible consequence of publishing open data is entirely new in our digital age. 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.

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. The idea of sharing information about paths to open data and interacting with others who understand the difficulty inherent in navigating the publication process of government data is vital to the communal atmosphere of the Open Data Alliance. While in other formal organizations it may feel challenging to voice opinions, this alliance should focus on welcoming all ideas and promoting openness within.

The standardization of data to ease the sharing.

While the informal sharing and combining of data among departments and jurisdictions are helpful in gaining insights, the biggest issue is the standardization of this data. The time and effort taken to clean this data for analysis purposes are of concern professionally and otherwise. It would be wise to utilize the variety of skills available in such an alliance to teach each other how cleaning and standardization can occur. Taking advantage of the skills and education of the group would be extremely beneficial to ease the sharing of data.

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

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. Like the standardization theme, the sharing of skills by teaching each other would create a support system that is reinforced by being able to be vulnerable enough to learn and strong enough to teach.