

Report of Contextual Design Research:
Identification of Information Needs for the City of Pittsburgh's
Data Governance Health Status Checks

DataTrackers, Group 5
LIS 2021: Identifying Information Needs
December 17, 2021

Table of Contents

Executive Summary	3
Background	3
Findings.....	3
Recommendations	4
Conclusion.....	4
Background	4
Methodology.....	5
Findings and Recommendations	6
Models	6
<i>Data in the Life Team Consolidated Model.....</i>	<i>6</i>
<i>Persona Model by RI</i>	<i>7</i>
<i>Identity Model with Matrices by LO</i>	<i>7</i>
<i>Collaboration Model by AW</i>	<i>8</i>
<i>Sequence Model by RW.....</i>	<i>8</i>
Affinity Diagram	9
Recommendations	9
Conclusion	11
Appendix A: Interview Protocol Sample Questions.....	12
Appendix B: Consolidated Models and Affinity Diagram	13
Appendix C: Key to the Matrices in the Identity Model with Matrices	14
References.....	15

Executive Summary

“For every minute spent organizing, an hour is earned.”

—Benjamin Franklin

“Start where you are. Use what you have. Do what you can.”

—Arthur Ashe

Background

In 2020, it was noted that the City of Pittsburgh was not in compliance with the Open Data Ordinance, which was enacted in 2014 and required all departments to conduct high-quality data management practices. To work toward compliance, the city’s Data Services Team from the Department of Innovation and Performance (I&P) planned data management engagement sessions to describe and discuss data management roles, responsibilities, and coordination.

For assistance in designing sustainable data governance practices with respect to the City of Pittsburgh’s Open Data Ordinance, the Data Services Team became a partner with the Library and Information Science department at the University of Pittsburgh School of Computing and Information. This meant a group of students enrolled in the Design Methods course sequence could be assigned to work with them to find and recommend solutions to their problem. Our group, DataTrackers, was assigned the task of working with the city to design Data Governance Health Status Checks to promote continuous updating of data inventories.

Findings

Through qualitative research involving interviews from a contextual design perspective, we gained significant insight into the ways city departments create, use, and manage their data and the difficulties they encounter. We also identified the following threats to consistent, ongoing data inventorying and open data publication:

- Poor data management practices due to the lack of resources
- Lack of policy governing the management of analog data
- Lack of policy and procedures surrounding the Open Data Ordinance
- Undefined roles and responsibilities with respect to open data publication
- Conflicts in reporting obligations
- Limited awareness of the collaborative role between I&P and city departments
- Lack of understanding of the requirements of the Open Data Ordinance

It was noted that the city departments did not feel that they were adequately qualified to manage data in compliance with the Open Data Ordinance. The departments also felt that their obligatory reporting to local, state, and federal agencies would take precedence over publishing open data. It also revealed interviewees identification as data users and lack of time or skills as the biggest threats to the Open Data Health Status Checks. Furthermore, interviewees admitted little knowledge of the ordinance itself and did not realize that much of the work they already perform is a step in providing data that would supply datasets for the open data ordinance.

Recommendations

The value of the qualitative research that we conducted is that the resulting solution(s) to the city's problem will be user centered. This is important because the city managers must take ownership of their data and their responsibilities with respect to open data publication. Solutions that take their current practices and struggles into consideration offer the best chance of reducing or eliminating their concerns as well as the existing threats to consistent, ongoing data inventorying and open data publication. A few plausible solutions are to

- Center Data Governance Health Status Checks around FOIL requests and data shared with other departments or agencies.
- Create a web application to assist users in keeping up-to-date with data inventories.
- Define procedures to remind and monitor progress.
- Develop submission guidelines and make them available as a web page and/or poster.
- Distribute a quarterly newsletter to highlight datasets that were published last quarter, projected datasets for the next quarter, and a department of the month.

Conclusion

We understand that some of the recommendations for Data Governance Health Status Checks mentioned in this report could be costly. We recommend that the organization consider starting with inventorying datasets that are already shared with other entities, i.e., data from government reports, FOIL requests, or interdepartmental requests, and implementing a technological solution to assist departments with understanding and completing them.

Background

The City of Pittsburgh Department of Innovation and Performance (I&P) provides IT support for all city departments and services. In addition to maintaining IT software and hardware, they also handle cybersecurity, digital projects, and department communications (including 311 calls and social media accounts). The Data Services Team in particular is responsible for providing data governance and management, including data analysis services and dashboard creation.

In 2014, the City of Pittsburgh enacted an open data ordinance, which required all departments to conduct high-quality data management practices, including the creation of an Open Data Catalogue; designating departmental Open Data Coordinators to update and maintain the catalogue; and submitting annual Open Data Reports.

While 'snapshots' of department datasets are openly available on the Western Pennsylvania Regional Data Center, I&P discovered in 2020 that the city is not in compliance with the 2014 Open Data Ordinance. To ensure compliance, the Data Services Team is planning data management engagement sessions and establishing roles and responsibilities for data coordinators and data stewards for city departments to undertake data governance.

Central to providing open data to the public is understanding what datasets are currently being utilized by city departments. Our group is charged with designing a sustainable process, called Data Governance Health Status Checks, that will promote continuous updating of data inventories. These

inventories will be, at first, simple spreadsheets that accurately list and describe all the data sets that are used by the various city departments. The process should include smooth check-in procedures so designated data stewards can update and report inventories to the Data Services Team.

To better understand how data is used and to discover problem areas with data management in city departments, the Data Services Team has asked that we interview managers of those departments who oversee data usage and data systems.

Methodology

Our role in this project was to conduct research that would inform the development of Data Governance Health Status Checks. The research goal was to identify a sustainable approach to maintaining data inventories and publishing open data, to formalize and standardize data governance for accountability and for reporting and measuring progress, and to identify threats to the success of these efforts. The design of the Data Governance Health Status Checks must take into consideration that the City of Pittsburgh is a multi-faceted organization with numerous departments that perform unique functions but that are interdependent in their use of data. Therefore, the research we conducted for this project employed a contextual design approach, which enabled us to gain insight into several department's practices and experiences around data use and management through a series of interviews designed for collecting relevant qualitative data. We consolidated the data and modeled the practices and experiences to reveal possible user-driven designs for the Data Governance Health Status Checks.

To ensure the validity and reliability of this research, we developed an interview protocol to make sure each interviewee was asked the same questions about data use, data management, and technology. See *Appendix A: Interview Protocol Sample Questions* for sample questions. To solicit candid responses, we informed each interviewee that their responses were confidential and that they could decline to answer any question. The interview protocol provided consistency and focus; however, we conducted each interview in a semi-structured format to create a relaxed atmosphere and to foster conversation that was partially directed by the user. This enabled us to gain deeper insights into each user's world. At least two people attended each interview with one person being the designated note taker. We held interpretation sessions to recall each interview and refine our notes.

To collect our data, we contacted several managers from a list provided to us by our partner contact. We interviewed seven managers from the following bureaus and departments in video conferences:

- Pittsburgh Water and Sewer Authority (PWSA)¹
- Pittsburgh Bureau of Fire
- Department of Mobility and Infrastructure (DOMI)
- Pittsburgh Bureau of Police
- Department of Public Works (DPW)

¹ PWSA is a separate entity from the City of Pittsburgh, i.e., it is not a department within the city. However, for the purposes of our models and analyses, we discuss PWSA in concert with other departments and refer to it informally as a city department.

- Office of the City Clerk

We organized the responses from each interview into separate files that we stored in our shared Google file system. To prevent readers from identifying interviewees from their responses, we anonymized each interviewee by assigning them a code, which consists of a letter to indicate the department and a number to indicate the interviewee. We used these codes instead of their names when we referred to them in models, diagram notes, and analyses.

From these interviews, we gained significant insight into the ways city departments create, use, and manage their data and the difficulties they encounter. Our team met weekly to discuss interviews, analyze the data, and build models. We built a team consolidated model, *Data in the Life*, to illustrate the lifecycle of data within City of Pittsburgh bureaus and departments. Each of us created an individual consolidated model to illustrate other patterns revealed by the data including personas, identities, collaborations, and work sequences. Finally, as a team, we created an affinity diagram to organize all the responses into one cohesive diagram to reveal concerns or problems common to all departments and interviewees.

Our team had numerous resources available to help us analyze the data and to create the models and affinity diagram. We used Zoom or Teams to conduct the interviews, to meet with our partner, and to meet as a team to work on the affinity diagram, team model, and corresponding analyses and presentations. We used the free versions of the online whiteboards MIRO and Mural or Microsoft Word to create the models and affinity diagram. Finally, we used a Google shared folder to organize our files.

Through our research, we have identified three primary threats to consistent, ongoing data inventorying and open data publication. We have also identified common practices and problems around data use and management. All of this allows us to recommend with confidence several possible solutions for formalizing and standardizing data governance into regular Data Governance Health Status Checks that takes into consideration the current data climate and incorporates appropriate steps to promote climate control.

Findings and Recommendations

Models

Data in the Life Team Consolidated Model

Our interviews revealed a series of challenges that could potentially hinder reliable and continuous updating of data inventorying and cataloging. We determined that there is inconsistent reporting and storage of data, so it can be difficult for departments to know what data they have; how to efficiently collect what they know they have; what data might be considered candidates for open platforms; and ultimately, what open data to report in inventories.

Inconsistent methods of data generation and management, and the variety of ways data is used, also have the potential to obscure inventorying because of the lack of standards. Part of the goal of data inventorying is to create a city-wide scheme that can be applied to the departments' disparate data sets that will support open data efforts, but the heterogenous quality of the datasets themselves can inhibit accurate and consistent reporting and identification of potential open data sets.

Other issues that could hinder consistent reporting of data inventories are conflicts with other reporting obligations. Our model shows the different outputs of data after they are collected on internal and external databases. Pittsburgh City departments are required by law to report various datasets in regular intervals to local, state, and federal governments. In addition, collaborative obligations with other city departments include sharing pertinent datasets. Our interviews have shown that these responsibilities take precedence over any open data efforts.

Persona Model by RI

A common pain point that was consistently noted in all the departments we interviewed was the dilemma in managing analog data also referred to as “paper entries.” Out of necessity, some data from the pre-digital world is still used in many of the city’s departments. This data is in the form of handwritten or typed documents, maps, charts, and other archival material.

The persona model highlights Richard Q. RecordKeeper. Richard is a composite of the four different viewpoints concerning analog data observed within the organization. 1) “I rely heavily on analog archival maps and charts to do my work” This mindset is common in departments where knowledge of the city’s layout is necessary. Sometimes this information is only available in its original paper format. 2) “I do not work at a desk; I rely on printed work orders to complete my assigned tasks.” This member of the organization does not have access to equipment needed to access digital data. 3) “We are not experts in records management, and we have no records management plan, no digitization, and no open data policy.” This organization member usually carries a heavy workload and does not have time or resources to learn current data handling methods. 4) “It would be great to have all this data available at the touch of a button, but I just don’t have time to add the information into a data management system.” While this member uses and understands data management tools, they would benefit from a designated data manager who would manage the department’s data. The ineffective methods noted produce challenges in creating consistent data management policies.

Identity Model with Matrices by LO

The *Identity Model with Matrices* includes a model of interviewees’ roles and experience with data and of what they like or think about data or data management. It also includes three matrices that compare employee data experience measures.

The *Identity Model with Matrices* reveals three primary challenges to overcome:

- Poor data management practices
- Limited or no skills or availability for open data practices
- Identification as a data user as opposed to a data manager

The “I Like, I Think” section reveals that interviewees like standardized, accessible, and consolidated systems. Although department O has significant problems because of poor data management and lack of related resources, the “My Role and Experience with Data” section reveals that most departments have some concerns about data accessibility, standardization, or consolidation. Any issues that hinder data acquisition and preparation will hinder compliance with the open data policy.

In addition to poor data management practices, many interviewees expressed concern that they did not have the ability or time necessary for open data publication. For example, interviewees V1 and M2 both have data savvy data analysts who are technologically capable. However, V1 believes that their

team does not have the skills to manage data or create dashboards. M2 says that their team has the ability but not the time because their time is spent preparing data for government reports.

The final challenge to overcome is that some interviewees identify as being ‘data users’ and **not** ‘data managers.’ Interviewees O1 and V1 have both identified very specifically as data users and have expressly stated that they are not data managers.

Collaboration Model by AW

The collaboration model analyzes the processes conducted between the Department of Innovation and Performance (I&P) and other city departments that were either discussed or overlooked during interview sessions. While many of I&P’s roles were covered, there were several variations and inconsistencies, especially relating to interviewees’ understanding of I&P’s role with open data.

I&P handles various communication processes for the city, including handling 311 calls from residents and forwarding them to appropriate city departments. Interviewees O1 and Z1, representatives from two departments that deal with 311 calls, mentioned I&P’s role in the service while other departments interviewed that also use that service did not. I&P also manages the social media content for all city departments, but only interviewees H1 and V1 acknowledged that role. Those same two interviewees mentioned that they held regular meetings with I&P. H1 talked about working with I&P to create an alert system, while V1 discussed how I&P assists with large data transfers.

The collaborative model also displays inconsistencies in awareness of I&P’s role in making departmental data openly accessible. While most departments acknowledged that I&P played some role in making some of their data publicly available, most did not know what that data was. Even the two departments who were aware of the type of data that was accessible could not specify details about that data. Another lacuna in the processes of this model is the lack of awareness of the Open Data Ordinance itself. In conclusion, the model suggests that there is limited awareness of the collaborative roles and responsibilities between I&P and city departments, and that current communication channels are minimal. A better understanding of interdepartmental roles and improved regular communication and training would be beneficial in promoting consistent data inventories.

Sequence Model by RW

The sequence model identifies the current processes of the various departments across the city of Pittsburgh. This model combines the practices of departments into a generalized process, but still identifies the unique practices of each department. A large sore spot that our group noticed was the lack of ubiquitous procedures and databases for each department. While many departments use the same programs and products, there are also large gaps when it comes to the digitization of data. For example, the V department enters their data manually into the system and multiple reports are filed whereas M utilizes a completely automated system. There also seems to be a hardship when it comes to knowledge of the open data ordinance. There also seems to be unique roadblocks in each department. For example, department V files the same event multiple times with separate records for the same event. This is quite confusing, and it seems as though there should be some work completed in every department to mainstream the data before sharing it with the public. In addition, many of the representatives we spoke with were not aware of any open data ordinance or requirements. I believe there should be a large push to firstly identify the legislation and make departments aware of the open data ordinance and goals for sharing the data to begin individual work within the departments to meet the standards necessary for sharing the data with the public.

Affinity Diagram

The affinity diagram process revealed several insights for our project. First, the City of Pittsburgh has a very diverse workforce with a diverse set of skills and perspectives. Although all managers support the Open Data Ordinance, they do not all believe that they have the resources, either time or skills, to comply with it. For example, O1 says that they oversee maintaining records, but that no one in their department is skilled in records management. Their data management practices are inconsistent, with their data stored on paper and six different digital storage systems. Another example involves M2 who says they have data analysts who could do this, but this team will prioritize government reporting over other data management tasks.

Lack of resources is not the only concern. Strong identification as a data user is another. V1 identifies strongly as a data user even though they are data savvy. V1 is afraid they will have a stack of requests sitting on their desk and that they will be responsible for noncompliance because they will focus on their problem-solving work, which is critically important and which they prefer. Identification as data users and lack of time or skills are the biggest threats to the Open Data Health Status checks.

Another insight is that several of the managers stated that they know very little or nothing about the Open Data Ordinance. Of the seven interviewed, O1, Z1, V1, and M2 said that they were not familiar with the ordinance. The City of Pittsburgh has not yet conducted the training for data coordinators and stewards. It is possible that some managers' concerns may be alleviated after they receive training. For example, V1 stated that they do not know how to create dashboards. However, the Data Services Team will provide support for these tasks if they know what data they have.

Finally, it is notable that every department shares data with other departments, with other government agencies, and/or with the public per the Freedom of Information Act legislation. These shared datasets are strong candidates for open data publication. When developing the Data Governance Health Status Checks, our group should consider both the concerns the managers have about resources and the work that they already do to prepare data for sharing.

Recommendations

The contextual design interviews revealed many pain points and ineffective methods that could threaten the city's efforts to comply with the Open Data Ordinance. The biggest threats to compliance are lack of good data management practices and identification as a data user, not a data manager. We recommend implementing one of the following three general solutions as a foundation for the specific implementation of Data Governance Health Status Checks.

- Hire a data manager to work with each department on their open data efforts, and eventually a team, to create policies and standards for data management and implement them throughout the organization.
- Train current managers on good data management practices.
- Implement non-invasive data governance² by formalizing current data practices, whether effective or not, and by noting areas for improvement.

The Data Governance Health Status Checks will formalize data governance around publishing open data and provide a system of support and accountability. We recommend including one or more of

² Seiner, Robert S. *Non-Invasive Data Governance*. First Edition. New Jersey: Technics Publications, LLC. 2014.

the following designs in the health status checks to support departments and to monitor progress with data inventorying and open data publication:

- Center Data Governance Health Status Checks around FOIL requests and data shared with other departments or agencies.
 - Define a set of expected datasets from past Freedom of Information Legislation (FOIL) requests, acquired from the Legal Department who receives and distributes requests to the departments. Consolidate similar datasets.
 - Prioritize datasets by frequency of the request and by frequency of the requesting organization, also acquired from the Legal Department.
 - Name data stewards and data coordinators based on who fills FOIL requests, also acquired from the Legal Department.
 - Define additional expected datasets by what departments report to the State and Federal governments, agencies, or other organizations.
 - Require data stewards to save all datasets from FOIL requests and government reporting. After distributing the request to the requesting individual or organization, data stewards will save the dataset to a shared drive. A standard filename format will be defined for consistency and the metadata will link the dataset to its data inventory entry.
- Create a web application to assist users in keeping up to date with data inventories. See San Francisco's data inventory submission form: <https://datasf.org/publishing/>.
 - A web form will act as a *data inventory and metadata wizard* for departments to enter information about their datasets.
 - This information will be stored in a database and can be filtered and exported to CSV files.
 - Data inventory entries can be viewed as web page reports and edited.
 - Data inventory entries can be copied and modified to reduce the time involved in entering records that are similar to existing records.
 - Whenever possible, data inventory fields will have a defined vocabulary, and users will be able to select an option from a list.
- Define procedures to remind and monitor progress.
 - Send regular reminder emails to complete data inventories.
 - Monitor data inventory and metadata entries, whether completed in the original Excel spreadsheet or through a web form. Someone from I&P will review the data inventory and metadata entries regularly and provide feedback.
- Develop submission guidelines and make them available as a web page and/or poster, similar to San Francisco's: <https://datasf.org/publishing/submission-guidelines/>.
 - Define data steward and data coordinator roles.

- Define steps for publishing open data: planning, identifying sensitive data, documenting data, etc.
- Offer incentives for publishing data. Distribute at meetings with all departments, VOLTRON or start new meetings around open data. Define the criteria and define the expected number of datasets for each department (based on raw count and percentage out of expected).
- Distribute a quarterly newsletter to highlight datasets that were published last quarter, projected datasets for the next quarter, and a department of the month (select based on raw number published, percentage out of expected, or growth/improvement over past quarters). Provide visualizations and tables of published datasets, expected datasets, missing inventories or missing fields or metadata from completed inventories.

Conclusion

A user-centered design for the Data Governance Health Status Checks is important for this project because the city managers must take ownership of their data and their responsibilities with respect to open data publication. The most impactful solutions will be those that take their current practices and struggles into consideration.

We understand that some of the recommendations for Data Governance Health Status Checks mentioned in this report could be costly. We recommend that the City of Pittsburgh consider starting with inventorying datasets that are already shared with other entities, i.e., data from government reports, FOIL requests, or interdepartmental requests, and implementing a technological solution to assist departments with understanding and completing data inventories. The following solutions have moderate to high impact potential and are achievable in the short term:

- Center Data Governance Health Status Checks around FOIL requests and data shared with other departments or agencies.
- Create a simple web application to assist users in keeping up to date with data inventories.
- Define procedures to remind and monitor progress.
- Develop submission guidelines and make them available as a web page and/or poster.
- Distribute a quarterly newsletter to highlight datasets that were published last quarter and projected datasets for the next quarter.

Appendix A: Interview Protocol Sample Questions

What is the primary business of your department?

What is your role with data in your department?

What are your top 5 datasets as far as importance to you or other people, i.e., other departments or the public?

What do you do with data in your department?

What do you know about the City of Pittsburgh's Open Data Legislation?

What is your perspective on the Open Data movement, in general?

What do you know about the proposed Data Governance Engagement Plan and Health Status Checks?

Describe the data you use in your work.

How do you handle data with personally identifiable information (PII)?

How do you collect or obtain your data? In what form is it when you get it, e.g., paper, computer?

What file formats have you worked with?

How much time do you spend working with data per day/week?

What criteria do you think must be met for data to be considered quality data?

How do you prepare/clean your data?

What are your documentation practices to describe your data?

How do you store your data?

In what ways do you work with other people on data tasks?

Do you report data to local, state, or federal agencies?

Do you share data with other city departments or the public?

Who are the stakeholders and users who access the data?

What do you think works well about how your department handles data?

What changes would you make to improve how your department handles data?

What are the IT systems and tools currently at your disposal?

What databases does your department use?

Appendix B: Consolidated Models and Affinity Diagram

As attachments to this report, please find the following models and the affinity diagram:

- *Data in the Life* Team Consolidated Model
- *Persona Model* by RI
- *Identity Model with Matrices** by LO
- *Collaboration Model* by AW
- *Sequence Model* by RW
- *Affinity Diagram* Team Assignment

*A key to the matrices in the *Identity Model with Matrices* appears in *Appendix C: Key to the Matrices in the Identity Model with Matrices*.

Appendix C: Key to the Matrices in the *Identity Model with Matrices*

Matrix A: Open Data Ability vs. Availability

- Has the required skills to prepare and publish open data
This metric was assessed from either a direct statement by the interviewee, as skilled or not, or an indirect statement that indicated they possessed some level of skill for preparing data for publication.
- Is available to prepare and publish open data
This metric was assessed by direct or indirect statements that the interviewee had the time to devote to open data publication, e.g., interviewee Z1 did not state directly that they had the ability or availability, however, data documentation tasks initiated by the interviewee indicate that they may be capable and available for such work as it relates to open data.

Matrix B: Open Data Ideas vs. Familiarity

- Has ideas and knowledge about what data could/should be open
This metric was assessed from a direct statement by the interviewee as having open datasets in mind.
- Is familiar with the Open Data Ordinance and understands the publication process
This metric was assessed from a direct statement by the interviewee but was further informed by other statements, such as V1's belief that they would be responsible for creating dashboards.

Matrix C: Data Reporting vs. Sharing

- Reports to the government or shares data with other city departments
This metric was assessed from a direct statement by the interviewee as having the responsibility to report to a government agency (local, state, or federal) or to fulfill requests for data from other city departments.
- Fulfills right-to-know requests or knows who does
This metric was assessed from a direct statement by the interviewee as receiving and fulfilling, or knowing who fulfills, requests from the public per the Freedom of Information Act (FOIL).

References

Seiner, Robert S. *Non-Invasive Data Governance*. First Edition. New Jersey: Technics Publications, LLC. 2014.