A2: Design and Implementation Plan
LIS 2022: Implementing Information Needs

City of Pittsburgh Data Services Team
Data Governance and Health Status Checks
Group 5, DataTrackers
February 13, 2022

Description of the Problem

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 Catalog; designating departmental Open Data Coordinators to update and maintain the catalog; and submitting annual Open Data Reports. The city of Pittsburgh's Data Services Team, which is part of the Department of Innovation and Performance (I&P), is responsible for supporting data governance and management for city departments. In 2020 the Data Services Team discovered that the City is not in compliance with the 2014 Open Data Ordinance and does not have procedures for prioritizing open data publication.

Our team is charged with designing a sustainable process, called Data Governance Health Status Checks, that will promote continuous updating of data inventories. Contextual inquiry interviews and modeling revealed several insights about how managers of city departments view data, their role with data, and the current processes in place for handling data. It is important to note that 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. Also, many managers identify strongly as data users and not as data managers. We will take these concerns into consideration and address them as best we can through our design.

Partner Feedback

Our group met with our partner, Chris Belasco, to discuss our final report and receive his feedback. We discussed focusing on Right-to-Know (RTK) requests, publishing newsletters to record and share progress reports, creating a website application for entering and updating data inventories and sharing datasets, and developing and defining data procedures that would include a way to track progress and a plan to keep the open data project visible to managers with regular reminders to complete data inventories. Chris agreed that the city of Pittsburgh should focus on publishing RTK requests and was pleased with the idea of a newsletter to report progress and keep the project visible. Chris also liked the idea of a website application, however, he explained that they want to focus on the process before putting time and money into a web application as it would be both expensive and time-consuming. Finally, Chris approved the idea of developing and defining procedures for keeping up with data inventories and tracking progress. He noted that it would be beneficial to track the progress that is being made, however, permission to track RTK requests would have to be approved by the mayor's office. Chris was very appreciative of our work because it revealed challenges that the City needs to address, such as the fact that many managers identify as data users and not data managers. Chris hopes that with clear procedures and guidelines, people's attitudes will change, because everyone manages data.

Description of Product and Scope

Our group proposes to create a set of *procedures* for Data Governance Health Status Checks, and if time permits, *user documentation* for the Data Governance Health Status Checks *procedures*. It is notable that every department shares data with other departments, with other government agencies, and/or with the public per the Right-to-know (RTK) law. These shared datasets are strong candidates for open data publication. We propose that the city of Pittsburgh take advantage of the work that managers already do to prepare data for sharing.

The *procedures* for Data Governance Health Status Checks will include a 1-3 page document that provides guidance on what to do after departments get a handle on the data they have. For example, the *procedures* will address what data to make open, how to prioritize datasets, how often to hold meetings, and how to track progress to ensure data managers are upholding their responsibilities. Tracking progress will involve collecting data—approximately 8 to 12 variables, such as the description and identification date of each identified dataset, the date the dataset is added to the data inventory, the date the dataset is determined to be open, or not, and the date the dataset is published, if applicable—to enable the Data Services team to measure each department's progress over time. The *procedures* will also include a template for a newsletter that will act as a progress report. The newsletter will be 1-3 pages in length and will include information about what each department has accomplished, what datasets they've made open, and what data inventories are in progress. It will also display aggregate data in tables and charts, i.e., counts and percentages for each department and city-wide totals, that are based on the progress indicators from the data. As an incentive, the newsletter will also include a section that highlights one department per newsletter for their work or growth in open data efforts.

The Data Governance Health Status Checks *procedures* will define a sustainable process to promote continuous updating of data inventories and provide users with the information they need to engage with the process. The *procedures* will also assist those managing data in prioritizing open data and maintaining some consistency across departments in relation to the open data process while allowing managers to maintain the data sharing process with which they are most familiar and comfortable.

Our focus will be on creating coherent and cohesive procedures, however, if time permits, we will create user-friendly content for their SharePoint website that provides instructions related to the *procedures* described above. For example, the information may include descriptions of the data coordinator and data steward roles and responsibilities, definitions of fields in the data inventory to allow side-by-side reference while filling out the inventory, and instructions of the workflow of the open data process, i.e., how to fill out the data inventory and keep it updated, where to store the data inventory, and a description of the lifecycle or routine of publishing open data. The *user documentation* may also include an infographic of the high-level process with information on where to get help for managers to display in their offices. The infographic will serve as both a resource and a reminder to work on data inventories and data publication. This poster will be a vector image to allow people to print it using their regular office printers or to enlarge it and have it professionally printed. The *user documentation* will provide clear instructions related to the Data Governance Health Status checks *procedures* and will keep the project visible to managers.

Additional Information

To ensure that our deliverables will cover the requirements set out by the Open Data Ordinance and expectations for health status checks, we will continue to have regular meetings with the Data Services Team. One key question we will need Data Services to address is whether I&P will have permission to track RTK datasets, as this will affect the content of our health status checks, as well as our documentation.

We will likely reach out to several of our previous interviewees for their feedback on our drafts of the procedures and user documentation to help guarantee usability and accessibility. In addition, as our group fine-tunes the steps involved for the user documentation, we might correspond with stakeholders involved with the process of making city data open. For example, a key partner is the Western Pennsylvania Regional Data Center, which helps facilitate accessibility of open data for the city of Pittsburgh. Their perspective will be necessary to help illustrate the data publishing process for users.

Scenario

Procedures and user documentation for City of Pittsburgh Data Managers

To ensure Pittsburgh Data Coordinators and Stewards make available consistent and accurate open data, our group is tasked with designing a set of procedures, and if time permits, corresponding user documentation.

Key:
- Data Coordinator (DC)
- Data Steward (DS)
- Department of Innovation and Performance (I&P)
- Western Pennsylvania Regional Data Center (WPRDC)

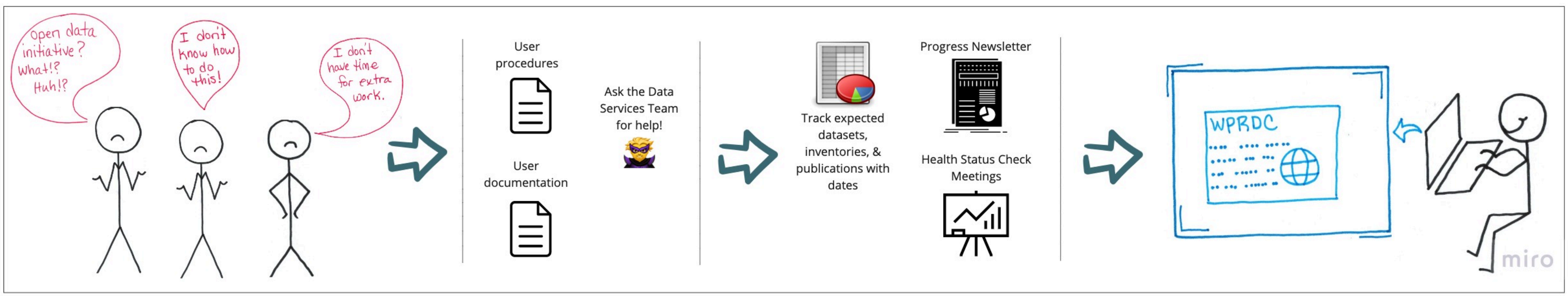
Key goals and needs
- Provide guidance for making data inventories with procedures and matching documentation.
- Ensure clear and accessible language is used to describe procedures.
- Discover problem areas that can be described in an accessible way.

Key pains and constraints
- Ensuring accessibility of procedures.
- Time constraints, which could inhibit learning and implementation.
- Data Services Team time constraints in assisting departments.

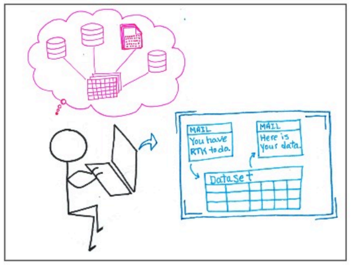
Context
- City of Pittsburgh departments will have assigned data coordinators and stewards who will upload data inventories.
- City Data Services Team will provide support.
- WPRDC makes data openly available.

Key activities and tasks
- Departments get a handle on their data (ideally RTK request data, first)
- Data Coordinators and Stewards access procedures and associated documentation
- Data inventory created and published to WPRDC.
- Keep up with meetings and communications with I&P.

Vision



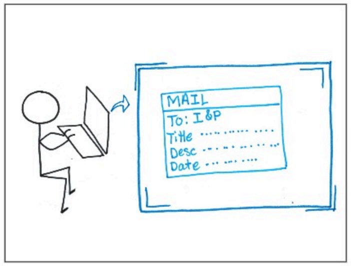
Storyboard



Step 1

Create or access a dataset for use and/or sharing among departments or other agencies, including RTK requests.

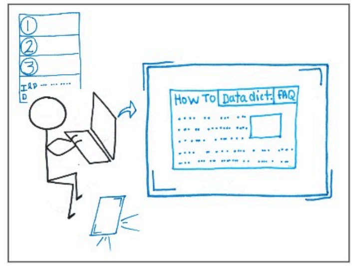
Who	DS
What	Dataset created or accessed from source
How	Use their regular process of creating and distributing datasets; contact DC and/or I&P for assistance



Step 2

Record initial tracking information about the dataset and share with the department of Innovation & Performance (I&P).

Who	DC
What	Tracking spreadsheet listing unique identifier, dataset title, description, date identified, data steward, and department
How	Email information or use a spreadsheet shared with I&P; reference procedures and user documentation on SharePoint; contact I&P for assistance



Step 3

Read procedures and call I&P for assistance.

Who	DC and I&P
What	Knowledge and confidence
How	Reference procedures on SharePoint

Step 4

Create the data inventory using the procedures for reference. Call I&P for assistance.

Who	DC and I&P
What	New record added to the Data inventory spreadsheet that accurately describes the data that flows in and out of a department
How	Reference procedures on SharePoint

Step 5

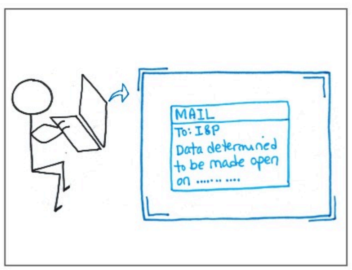
Record that the data inventory was completed and share the date with I&P.

Who	DC
What	Tracking spreadsheet with the date that the inventory was completed for given dataset
How	Email information or use a spreadsheet shared with I&P; reference procedures and user documentation on SharePoint; contact I&P for assistance

Step 6

Determine if the dataset should be made openly available to the public.

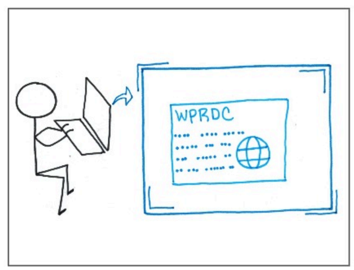
Who	DC
What	Openness designation for given dataset
How	Reference data inventory and procedures on SharePoint



Step 7

Record the openness determination and share the date with I&P.

Who	DC
What	Tracking spreadsheet with the date that the openness determination was made for given dataset
How	Email information or use a spreadsheet shared with I&P; reference procedures and user documentation on SharePoint; contact I&P for assistance



Step 8

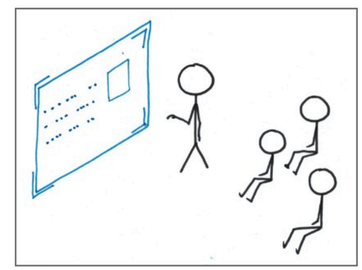
Publish the data on the WPRDC portal.

Who	DS and I&P and/or WPRDC
What	WPRDC provides the technical infrastructure to host and make available open data sets
How	Submit datasets in CSV format to WPRDC using process specified in procedures

Step 9

Record that data is published and share the date with I&P.

Who	DC
What	Tracking spreadsheet with date published for given dataset
How	Email information or use a spreadsheet shared with I&P; reference procedures and user documentation on SharePoint; contact I&P for assistance



Step 10

Participate in periodic check-ins with I&P

Who	DS, DC, I&P
What	Annual status check where Departments review status of data management activities with Data Governance Committee
How	Schedule meetings with stakeholders. Ensure necessary data management is ready to share

Prototype Plan

What are the salient features of your implementation you want to test with a prototype?

The feature to be tested includes the multistep process for making data open, which is portrayed in the attached storyboard. The steps for making data open, along with questions that we need to answer to create the prototype, follow:

1. Create or access a dataset for use and/or sharing among departments or other agencies, including RTK requests.

   This step includes fulfillment of any data sharing request so workflows that are familiar to data stewards are not interrupted.

2. Record initial tracking information about the dataset and share with the department of Innovation & Performance (I&P).

   Step 2 is the first step in the tracking process. The data coordinator (DC) will send the following information to I&P: the unique identifier, title, description, and identification date of the dataset and the name of the data steward and department. The status of this dataset will be 'inventory expected.'

3. Read procedures and call I&P for assistance with the data inventory.

   This step is repeatable and can be visited at any point. We need to determine where procedures will be made available to users, who users should contact for help, and how they should contact that person (phone, email).

4. Create the data inventory using the procedures for reference. Call I&P for assistance if necessary.

   We need to review the current documentation for creating data inventories and determine if any additional information or modifications would be helpful to the users. We also need to find out how each department can share their data inventory with I&P.

5. Record that the data inventory was completed and share the date with I&P.

   Step 5 is the second step in the tracking process. The data coordinator (DC) will send the following information to I&P: the unique identifier and inventory date. The status of this dataset will be 'inventory completed.'

6. Determine if the dataset should be made openly available to the public.

   Use the data inventory to determine if the dataset should be made open. Dataset prioritization involves evaluating the sensitivity level of the data, the anonymity of the data or the ability of the data to be anonymized, and the value of the dataset to stakeholders.

   Note: All RTK data is priority because organizations are required by law to fill these requests for information. As soon as data is distributed to the requesting agency, it belongs to everyone and should be published to the open data portal.

7. Record the openness determination and share the date with I&P.

   Step 7 is the third step in the tracking process. The data coordinator (DC) will send the following information to I&P: the unique identifier and openness determination date. The status of this dataset will be 'expected to be made open.'

8. Publish the data on the WPRDC portal.

   This step involves determining how data will be 'moved' to publication. We discussed the following three options for streamlining the publication process:

   a. CSV file upload to WPRDC
   b. Cloud data warehouse - application to table pipeline to get data from secure place
   c. Secure file transfer protocol - WPRDC pulls via SFTP

9. Record that data is published and share the date with I&P.

   Step 9 is the fourth step in the tracking process. The data coordinator (DC) will send the following information to I&P: the unique identifier and publication date. The status of this dataset will be 'published.'

10. Participate in periodic check-ins with I&P.

    We need to research a reasonable interval for holding meetings, i.e., quarterly, tri- or bi-annually. We also need to create a mockup for the newsletter with a section that highlights departments' progress.

The tracking process allows the city to measure their progress and to hold data stewards and coordinators accountable for identifying datasets, creating inventories, determining whether to make data open, and publishing data. We are still uncertain whether I&P will be able to track datasets from the point of identification or if they will have to wait until the data is added to the data inventory before they can track it. We believe they will at least be able to track data at the point that the data is inventoried, which will eliminate Step 2. This will make it difficult to measure progress, because knowing when a dataset has been identified provides a more complete picture of what data departments have and how productive they are in completing the process. However, the process will still be more helpful than no process at all even if we must eliminate Step 2.

What questions do you want to answer with your prototype?

- What questions do the users have for us or I&P as they go through our procedures?
- Is I&P spending too much time supporting the procedures?
- Are users asking questions that are very specific to their issue that do not affect anyone else, or are they asking questions that others may have?
- Do users say that anything is confusing or unclear?
- Considering answers to the above questions, should the procedures be revised?

What prototyping method do you want to use?

We plan to produce a high-fidelity prototype to represent our design solution that will take the form of a draft of our procedures. A high-fidelity prototype will ensure that the procedures and corresponding document features and formatting will be assessed for accuracy and usability.

What kind of users will be exposed to your prototype?

In addition to our contact for our partner organization, we hope to share our prototype document with several data managers from various city departments for their thoughts and feedback. These users will likely not only use our procedures but will have a good idea of the expectations of other department users as well.

How will you collect feedback from the users?

Feedback will likely be collected from our partner contact and city stakeholders through email. Depending on user availability, Zoom calls might also be used to collect user thoughts and feedback.

Implementation Timeline

Group Meetings	Goals	Assignment	Due Date
2/15/2022	**Meet with partner to clarify questions and expectations about the prototype**		
2/22/2022	**Draft prototype of procedures**		
	Procedure Steps (All members)		
	Research how to prioritize making data open (Robyne)		
	How often to hold meetings (Reyann)		
	How to track progress (Lisa)		
	Define incentives / Outline Newletter (Andrew)		
2/27/2022	**Prototype testing and feedback**	Prototype	3/4/2022
3/8/2022	Break		
3/15/2022	Gather feedback - emails and interviews with users involved in testing		
3/22/2022	Create feedback grid	Feedback Grid	3/27/2022
3/29/2022	Work on final design - incorporate feedback, add to or clarify procedures, if necessary		
4/5/2022	Work on final design - create user documentation, if time permits		
4/12/2022	Complete design - refine language and format		
4/19/2022	Final presentation	Final Product	4/24/2022