A3: Prototype Report
LIS 2022: Implementing Information Needs

City of Pittsburgh Data Services Team
Data Governance and Health Status Checks
Group 5, DataTrackers
March 4, 2022

Introduction

In 2014, the city of Pittsburgh enacted an Open Data Ordinance (ODO), 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 designing a sustainable process, called Data Governance Health Status Checks, that will promote continuous updating of data inventories. Due to receiving new information from the Data Services Team after we submitted our Implementation Plan, we made some last-minute changes to make sure we addressed the city's information needs. Our original plan was to document procedures for both the Data Services Team and the data coordinators, however, some of the procedures that we outlined in our storyboard became obsolete with the new information we received from our partner. Our prototype still includes documentation for both the Data Services Team and data coordinators, but it focuses more on the procedures to be carried out by the Data Services Team. Ultimately, this will benefit managers of all departments, because the Data Services Team will be in a better place to support them in maintaining their data and data inventories. There are three major components to our product:

Health Status Checks Procedures

This component consists of the procedural documentation surrounding Health Status Checks.

- *Data Coordinator Procedures* provides data coordinators with basic instructions for completing the data inventory and submitting the data inventory and datasets ready for publication to the Data Services Team.

- *Prioritizing Data Publication* provides guidance for selecting datasets for publication based on several features including how useful or popular the dataset is.

- *Data Services Procedures* provides the Data Services Team with instructions for creating and maintaining various tables that we designed for the Health Status Checks and a newsletter, which will be the primary reference document used in Data Governance Health Status Checks meetings.

- *Health Status Checks Newsletter Template* provides an outline for the Data Services Team to report progress to departments and other stakeholders.

Tracking and Summary Tables

This component consists of a set of tables in Excel that will allow the Data Services Team to keep track of milestones for each department, such as whether the department has scheduled or completed a Needs Assessment, has completed their data inventory and data catalog, and has identified a data coordinator and data stewards. All of these data points ultimately allow the Data Services Team to track which departments are compliant with the ODO and take action to assist noncompliant departments in becoming compliant. This component also consists of a user guide with detailed instructions for entering data into the tracking table and understanding the summary tables, which update automatically when the tracking data is updated.

Critical Records Program

This component is a small visual basic application in Excel that searches data inventory files for datasets that are due to be updated, that should be open but are not yet open, and/or that have errors in fields used to identify these problems. The program creates a report that classifies these problems into one of three criticality levels—Mission Critical, Vital, or Important, which are displayed in separate sheets in the report. This report will promote continuous updating of data inventories by highlighting datasets that need to be reviewed for updates or that have other problems that could result in data loss. This component also consists of a user guide with detailed instructions for running the program and copying data into the Newsletter.

File Map

Below is a map of the files in the submitted folder.

DataTrackers Prototype (root folder)

1. Prototype report (.docx)
2. Data Inventory Template (.xlsx)
3. Health Status Checks Procedures (folder)
    a. Data Services Procedures (.docx)
    b. Health Status Checks Newsletter Template (.pptx)
    c. Data Coordinator Procedures (.docx)
    d. Prioritizing Data Publication (.docx)
4. Tracking and Summary (folder)
    a. Tracking and Summary Tables (.xlsx)

    b. User Guide: Tracking and Summary Tables (.docx)
5. Critical Records Program (folder)
    a. Critical Records Report Macro (.xlsm)
    b. User Guide: Critical Records Program (.docx)
    c. Data Inventory Files (folder) - contains mock data inventory files for testing (*.xlsx)
    d. Critical Records Reports (folder) - contains Critical Records Program output files (*.xlsx)

Audience and Testing

There are two audiences for this product: 1) the Department of Innovation and Performance, specifically, the Data Services Team, who oversee the city's efforts to publish open data, and 2) the data coordinators from each city department who maintain data inventories and publish their department's data. The following table explains the audience and the testing we will conduct for each of the deliverables in this product.

<table>
<tr><td>Data Services Procedures.docx</td><td>Audience: Data Services Team</td></tr>
<tr><td colspan="2">Usability Testing: Members of the Data Services Team will carry out the steps in the procedures and provide feedback about their experiences. We are specifically looking to find out if they need help performing any steps or if they have questions about the process while they are going through it. Their feedback will help us to improve the document by clarifying or adding information.</td></tr>
<tr><td>Health Status Checks Newsletter Template.pptx</td><td>Audience: Data Services and Data Coordinators</td></tr>
<tr><td colspan="2">Design Critique: Members of the Data Services Team and two to four Data Coordinators will review the design of the newsletter and provide feedback about what aspects of the design meet their needs and which aspects need some improvement. We are specifically looking to find out how the layout and content of the newsletter contribute to its feasibility and usefulness.</td></tr>
<tr><td>Data Coordinator Procedures.docx</td><td>Audience: Data Coordinators</td></tr>
<tr><td colspan="2">Usability Testing: Two to four Data Coordinators from different departments will carry out the steps in the procedures and provide feedback about their experiences. We are specifically looking to find out if they need help performing any steps or if they have questions about the process while they are going through it. Their feedback will help us to improve the document by clarifying or adding information.</td></tr>
<tr><td>Prioritizing Data Publication.docx</td><td>Audience: Data Services and Data Coordinators</td></tr>
<tr><td colspan="2">Design Critique: Members of the Data Services Team will review the guidelines for prioritizing data publication and provide feedback about whether the guidelines match their expectations for how departments should prioritize data publication. After incorporating comments from the Data Services Team, two to four Data Coordinators will review the guidelines for coherence and clarity.</td></tr>
</table>

User Guide: Tracking and Summary Tables.docx	**Audience:** Data Services
Usability Testing: Members of the Data Services Team will review the user guide for the Tracking and Summary Tables process, carry out the steps to update the Data table, and provide feedback about their experiences. We are specifically looking to find out if the User Guide is clear about which tables they are supposed to maintain and which tables get copied to the newsletter. Their feedback will help us to improve the document by clarifying or adding information.	
Tracking and Summary Tables	**Audience:** Data Services
Focus Group: Members of the Data Services Team will review the summary tables and discuss the indicators. We are specifically looking to find out if the indicators meet their needs for monitoring the open data process or if they would like to modify, remove, or add any. We also need to verify which fields are used to determine compliance with the ODO.	
User Guide: Critical Records Program.docx	**Audience:** Data Services
Usability Testing: Members of the Data Services Team will review the user guide for the Critical Records Program, carry out the steps to prepare and run the macro, and provide feedback about their experiences. We are specifically looking to find out if the user guide is clear about how to change the paths to the source and target folders and how to run the program. Their feedback will help us to improve the document by clarifying or adding information.	
Critical Records Program	**Audience:** Data Services
Focus Group: Members of the Data Services Team will review the summary tables and discuss the types of problems caught by the program and the distribution of those problems into the three levels of criticality. We are specifically looking to find out if the program will address their concerns about losing track of datasets in the data inventories. Their feedback will help us to improve the program by modifying its problem and/or levels of criticality definitions.	

Information to Gather and Questions for Testing

In order to evaluate the effectiveness of the Health Status Checks we have developed questions for those testing the plan. The following questions will help us determine what part of our plan the testers found most useful as well as their suggestions for improvement. We recommended that testers read the Health Status Checks document in its entirety before reviewing the questions.

Data Inventory

1. Did you find Data Validation columns helpful?
    a. Data validation columns can measure for other criteria such as the sensitivity and quality ratings columns

    b. The Critical Records Program performs data validation so the data inventory does not have to implement this, however, the program could report too many errors (typos) if the data inventory template does not also implement it.
2. Would it be beneficial to have instructions for creating and linking to data validation tables for future editing?
3. Would it be beneficial to have a ‘map’ of the Excel formulas in the Tracking and Summary Overall and Department sheets?

Tracking and Summary Tables

1. In the Tracking and Summary Tables.xlsx ‘Data’ sheet, would you like to maintain the Data Stewards table (‘Data Stewards’ sheet), which the Calculated_Count_Stewards column uses to calculate the count of data stewards, or would you rather use the Count_Stewards column and enter the count of Data Stewards yourself?
2. The table in the Department sheet has more columns than you suggested. You wanted the number of stewards per department. All other columns are extra. Please let us know if you want to change these - add others or remove any.
3. What is required for a department to be considered compliant?

Critical Records Program

1. Do the criticality levels work for you? Should any be redefined?
2. Do you want any additional fields displayed in the output?
3. Do you want a date range check? For example, if an update date is inadvertently set to a future date, the entry will not be flagged as due or overdue at the right time.
4. Do you want to catch restricted and private datasets that are marked as open? This would most likely be an error in the data inventory and not that the restricted/private dataset is open when it shouldn’t be.

Newsletter

1. How many aspects of the newsletter can be reasonably carried out by Data Servers staff?
2. What are the criteria for selecting datasets to be listed in the newsletter?
3. What are the criteria for selecting a featured department? What are the benchmarks for exceptional open data practices?
4. Who are the users for the newsletters?
5. How should Data Services staff share the newsletter with designated users?

Expected Changes

Our meetings with the Data Services Team have helped us to narrow the focus of the process that we are tasked to design, as well as to specify the particular user groups who will rely on our documents. After these important meetings, we are confident that our procedures will not require major overhauls going forward. However, there are a number of modifications with all the documents in our prototype that we are interested in implementing and testing with our partner organization. These expected changes will help fine-tune our procedures to ensure usability and utility for data services staff and city department data managers.

Health Status Checks Procedures

1. We would like to add a section in both Health Status Check Procedures documents that briefly explain the Open Data Ordinance and what it means to be in compliance. We think this will help keep the processes described in those documents grounded in the city's open data objectives.

Data Inventory

1. We implemented data validation for some fields in the Data Inventory template to make sure users enter valid data. We believe this will also assist users in entering the data as they will select options from a drop-down menu instead of typing the values. The Data Services Team may decide not to include this feature.

Tracking and Summary Tables

1. The department summary table, which captures aggregate information for each department's open data efforts, will likely change or expand depending on the feedback from the Data Services Team.
2. While we endeavored to include as many fields as we felt were necessary to compute whether or not a department is compliant, we expect a more refined definition of compliance from our partner so these fields will likely change.

Critical Records Program

1. The range of critical level values may look different; in particular, time-related criteria may change (such as considered Due Soon, Due, and Overdue).
2. We are also interested in limiting the data range that is possible for users to input.
3. There might be additional criteria for the types of problems found, so not only will there potentially be more variables, but there might also have to be corresponding changes to the program that runs the reports.
4. Additional output fields from data inventories might also be considered to include in the critical reports.

Newsletter

1. We would like to have a small program that prints the first page of the newsletter and the tables to one PDF file, so the user does not have to copy and paste tables into the newsletter.

We are still working on a solution for this, but we will have to confirm our other questions about the newsletter before devoting too much time to writing a program.

2. Formatting is another aspect of the newsletter that will likely change. This will likely be done by Data Services Team themselves, but we will at the very least ensure that all necessary content is accounted for in the newsletter.