A8: Final Report
LIS 2022: Implementing Information Needs

City of Pittsburgh Data Services Team
Data Governance and Health Status Checks
Group 5, DataTrackers
April 24, 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 was 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 took these concerns into consideration and designed a process and tools that would assist the Data Services Team in supporting data coordinators in their efforts to comply with the Open Data Ordinance.

The scope of our project changed since the Design and Implementation Plan. 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. In addition, our partner contact gave us a set of data points they wanted to use as progress indicators and expressed concern about data loss from the data inventory, i.e., losing track of datasets that were recorded but past due for updates. He was afraid of data updates being missed or late. We used this new information along with the information we gathered during the contextual design interviews to design a product that includes procedures for both the Data Services Team and data coordinators as well as tools for tracking and reporting progress and for discovering data inventory records that have missing, invalid, or illogical data or that are due to be updated. How and when to use these tools are included as steps in the procedures, and each tool has an accompanying user guide with more information about its use and maintenance.

The Final Design

In response to the Data Services Team request for assistance, DataTrackers created a product that will assist the city of Pittsburgh's Data Services Team as well as city departments in meeting compliance guidelines. DataTrackers produced a Visual Basic macro, a newsletter template and accompanying procedures for Data Services and Data Coordinators. These products will address the problem of non-compliance by creating an easy method for departments to evaluate and publish a data inventory as it pertains to the city's Open Data Ordinance.

Other deliverables DataTrackers considered for production were procedures outlining the publication of open data on the Western PA Regional Data Center (WPRDC). We decided against that because the Data Services Team assumes that responsibility and did not need guidance for that task. DataTrackers also

considered creating a website application for entering and updating data inventories and sharing datasets. However, we discovered during our meetings that the Data Services Team already has a SharePoint site where they house this information, and a website would be redundant.

Our final product consists of the following major components:

Data Services Procedures

Documentation providing step-by-step instructions for conducting Data Governance Health Status Checks. Users of this guide should have a basic understanding of Pittsburgh's Open Data Ordinance. These procedures provide a detailed breakdown of the steps involved in tracking progress, identifying data inventory records that need attention, and conducting Health Status Checks meetings.

Tracking and Summary Tables

A set of Excel tables for tracking and reporting city-wide and departmental progress toward compliance with the Open Data Ordinance. This component is accompanied by user documentation for the Data Services Team: *User Guide: Tracking and Summary Tables*. These tables are updated by the Data Services Team and shared with department data coordinators in the newsletter.

Critical Records Program

The Critical Records Program is a Visual Basic macro designed to prevent data loss from data inventories. The macro discovers data inventory records that are overdue, due, or due soon for updates or that have missing, invalid, or illogical data. Reports will be shared by the Data Services Team at Data Governance Health Status Checks meetings. Basic computer skills are all that is required to run the program. Familiarity with Visual Basic for Applications or a reasonable comfort level with programming in general is required to modify the program. This component is accompanied by user documentation for the Data Services Team: *User Guide: Critical Records Program*.

Data Coordinator Procedures

Documentation for data coordinators that describes the steps involved with preparing for and participating in Data Governance Health Status Checks. This document is a summary of the major steps necessary for data coordinators to complete and update data inventories and follow good data management practices, which are required for compliance with the Open Data Ordinance.

Prioritizing Data Publication

Documentation for data coordinators to assist them in establishing initial priorities for selecting datasets for data inventory as well as for open data publication.

Health Status Checks Newsletter Template

A PowerPoint file with space for reporting progress to data coordinators and other stakeholders. The file can be edited as needed. This component is accompanied by user documentation for data coordinators: *User Guide: Newsletter*. The document provides an overview of the newsletter structure and describes the four major sections: 'News,' 'Featured Department,' 'Summary Tables,' and 'Critical Records Tables.' The subsequent components of the guide go into further detail for each newsletter section. A description for how Data Services selects a 'Featured Department' is also provided.

All documents listed above—*Data Services Procedures*, *Data Coordinator Procedures*, and *Prioritizing Data Publication*—and the user guides that accompany the other components are included in the *Data Governance Health Status Checks Handbook*, which is organized as follows:
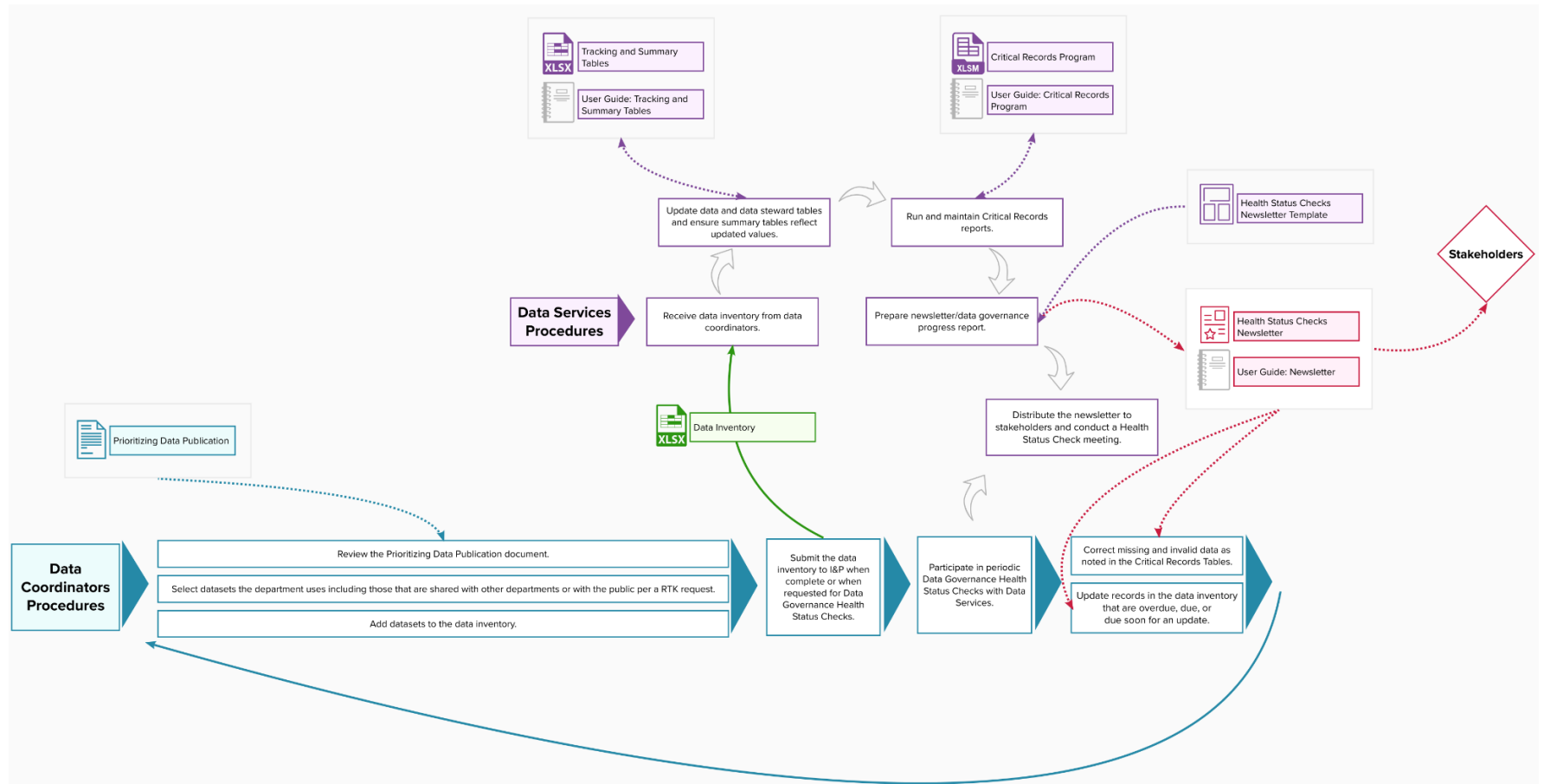
- Data Services Procedures

- User Guide: Tracking and Summary Tables
- User Guide: Critical Records Reports
- Data Coordinator Artifacts:
    - Data Coordinator Procedures
    - Prioritizing Data Publication
    - User Guide: Newsletter

Transition Plan

Our vision for how Data Services uses our product is best illustrated as a lifecycle. This *Data Governance Lifecycle* starts with data coordinators completing their data inventories. The process is not linear as it is portrayed in the written procedures. The first three steps coincide as data coordinators review the *Prioritizing Data Publication* document, select datasets, and add datasets to their data inventories concurrently. When data coordinators submit their completed data inventories, Data Services prepares for the Data Governance Health Status Checks by updating the *Tracking and Summary Tables* and running the *Critical Records Program*. Data Services then creates the newsletter, which acts as a data governance progress report for stakeholders, and conducts the Data Governance Health Status Checks meeting. Data coordinators review the Critical Records Tables in the newsletter and start work to address the health status notifications. Finally, the process starts over as data coordinators continue to add datasets to their data inventories.

Data Governance Health Status Checks Lifecycle

Tracking and Summary Tables (XLSX)

User Guide: Tracking and Summary Tables

Critical Records Program (XLSM)

User Guide: Critical Records Program

Health Status Checks Newsletter Template

Health Status Checks Newsletter

User Guide: Newsletter

Update data and data steward tables and ensure summary tables reflect updated values.

Run and maintain Critical Records reports.

Stakeholders

Data Services Procedures

Receive data inventory from data coordinators.

Prepare newsletter/data governance progress report.

Distribute the newsletter to stakeholders and conduct a Health Status Check meeting.

Prioritizing Data Publication

Data Inventory (XLSX)

Data Coordinators Procedures

Review the Prioritizing Data Publication document.

Select datasets the department uses including those that are shared with other departments or with the public per a RTK request.

Add datasets to the data inventory.

Submit the data inventory to I&P when complete or when requested for Data Governance Health Status Checks.

Participate in periodic Data Governance Health Status Checks with Data Services.

Correct missing and invalid data as noted in the Critical Records Tables.

Update records in the data inventory that are overdue, due, or due soon for an update.

Data Services Procedures

The first major component of our documentation is the *Data Services Procedures*. This document provides a high-level overview of the five major steps necessary for the Data Services team to carry out Data Governance Health Status Checks. The document begins with a brief overview of its scope and a description of the checks followed by detailed descriptions of or instructions for each step. Users from the Data Services team should at least have a basic understanding of Pittsburgh's Open Data Ordinance, as well as a familiarity with related Data Services open data documentation and engagement sessions to perform these procedures.

Tracking and Summary Tables

The *Tracking and Summary Tables* represent a snapshot of the city's overall compliance with the Open Data Ordinance and will be updated by the Data Services Team and shared with department data coordinators in the newsletter. Data Services should consult this guide when inputting and updating tracking and summary data and data steward information. Users from the Data Services Team should at least have a basic understanding of Pittsburgh's Open Data Ordinance, as well as a familiarity with related Data Services open data documentation and engagement sessions. A basic grasp of Microsoft Excel is also helpful for understanding the summary calculations. Overall, the guide should deliver the necessary support for Data Services to update tracking and summary data.

The *User Guide: Tracking and Summary Tables* begins with a brief description of the purpose and scope of the tables and the guide itself. The guide also provides a table that describes the four primary sheets in the Tracking and Summary.xlsx file: 'City,' 'Department,' 'Data,' and 'Data Stewards.' The 'City' and 'Department' sheets are composed of variables that automatically summarize the information in the 'Data' and 'Data Stewards' sheets.

The next two sections of the guide include instructions for updating the two sheets which require manual data entry by the Data Services Team: 'Data' and 'Data Stewards.' The section for updating the 'Data' sheet includes a table that defines each variable in the sheet and designates the variables as a particular Type: *user populated text*, *user populated count*, and *calculated*, which are also defined. The final section of this guide is a brief description of the 'Data Stewards' sheet, and instructions for updating the variables.

Critical Records Program

Concern regarding data loss from the data inventories inspired the *Critical Records Program* component of this product. The large number of expected datasets with different update dates and update frequencies will be very difficult to maintain manually. The *Critical Records Program* is designed to alleviate this concern by discovering data inventory records that are overdue, due, or due soon for updates or that have missing, invalid, or illogical data.

The *Critical Records Program* reads data inventory records from multiple data inventory files to discover and write *critical* records to a report. Critical records are records that have one or more of the following health status notifications:

- Records are 'Public' but 'Not yet open'
- Restricted or Private records that do not have a 'Public' record in the data inventory
- Records are overdue for updates
- Records are due for updates
- Records should be updated soon

- Missing or invalid data in one or more of the following critical fields:
    - Date Created/Imported
    - Internal Priority
    - Update Frequency
    - Date Updated
    - Data Classification
    - Sensitivity Rating
    - Public Status

Invalid values include any values that do not match one of the values accepted for a given field or that are not dates in a date field, and this includes values with minor typographical errors or inconsistent capitalization. The *Critical Records Program* assigns a health status to each critical record to alert users about what needs to be done to maintain the data inventory record and resolve the health status notifications for it.

The Data Services Team will run the *Critical Records Program* before each Data Governance Health Status Checks meeting. They will be able to compare the current report to past reports to verify that past health status notifications were resolved, and they will distribute new or unresolved health status notifications to data coordinators for resolution.

The *User Guide: Critical Records Program* provides instructions for running and modifying certain aspects of the program. Basic computer skills are all that is required to run the program. The only necessary change to the code is the source path and target path constants, which need to be values that reference the user's computer or network. Other modifications require familiarity with Visual Basic for Applications or a reasonable comfort level with programming in general, but these modifications are not required to run and benefit from the program. The program code is also extensively documented for reference. The program code is extensively documented for reference.

The user guide provides the following information about the program:

- **Data Inventory** This section describes the features of the data inventory that were implemented both to assist users in data entry and to minimize the number of critical records that arise due to invalid data. It also describes the protocol for naming datasets and explains how the program references data inventory columns in case these specifications need to be changed.

- **Prepare to Run the Critical Records Program** This section describes the steps that need to be taken or considered before running the program.

- **Run the Critical Records Program** This section provides steps for running the program and accessing the critical records report.

- **Due Dates, Update Windows, and Level of Criticality Assignment** This section explains how the program calculates the date update is due and the update due windows, which are used to distinguish between records that are due versus due soon for updating. It also describes the level of criticality assignments and what data inventory record characteristics are used to assign a record to a particular level. Finally, this section provides instructions for modifying how the program calculates the update due date, update due windows, and level of criticality assignment.

- **Modify the Program Code** This section provides steps for opening the Visual Basic Editor and outlines the changes that might be necessary. This step requires familiarity with Visual Basic for Applications or a reasonable comfort level with programming.

Data Coordinator Procedures

The *Data Coordinator Procedures* is a summary of the major steps necessary for City of Pittsburgh department data coordinators to prepare for and participate in Data Governance Health Status Checks. These procedures outline what each Pittsburgh department must do to comply with the city's Open Data Ordinance and to follow good data management practices by completing a data inventory of all datasets used by the department. Meaningful engagement with these procedures requires data coordinators to first have a good understanding of the Open Data Ordinance and to attend the City of Pittsburgh Data Governance Engagement Sessions. The *Data Coordinator Procedures* is critical to the success of the Data Governance Health Status Checks and Open Data Ordinance project. This document should be updated to reflect changes in the project with respect to completing and submitting data inventories or to resolve any confusion that might arise when data coordinators start to use it. There are no technical functions or specifications to maintain or update this document, so there are no transitional instructions to provide other than to review and update this document as needed.

Prioritizing Data Publication

The *Prioritizing Data Publication* document will assist data coordinators in establishing initial priorities for selecting datasets for the data inventory and for open data publication. To benefit from the guidance this document provides, data coordinators must first have a good understanding of the Open Data Ordinance and attend the City of Pittsburgh Data Governance Engagement Sessions. The *Prioritizing Data Publication* is the first resource data coordinators are given to help them determine which datasets are high priority and which are low priority with respect to adding them to the data inventory. This document should be updated to reflect changes in the project with respect to prioritizing datasets or to resolve any confusion that might arise when data coordinators start to use it. There are no technical functions or specifications to maintain or update this document, so there are no transitional instructions to provide other than to review and update this document as needed. These guidelines were developed in accordance with the open data guidelines from the Sunlight Foundation.

Health Status Checks Newsletter Template

Data Services will distribute newsletters to department data coordinators on a regular basis that feature news relating to Pittsburgh's open data efforts, as well as department and city-wide progress towards compliance. The *User Guide: Newsletter* includes descriptions for each segment of the newsletter and its associated tables. Like the other user guides, a basic understanding of the Open Data Ordinance is required to use this guide.

The document begins with an overview of the newsletter structure, describing the four major sections: 'News,' 'Featured Department,' 'Summary Tables,' and 'Critical Records Tables.' The subsequent components of the guide go into further detail for each newsletter section. The 'News and Featured Department' component provides descriptions for the type of information that is highlighted in the first page of the newsletter, such as announcements and events regarding Open Data in the city. A description for how Data Services should select a 'Featured Department' are also provided.

The 'Summary Tables' and 'Critical Records Tables' components of the guide describe the scope, purpose, and variables for each table to assist data coordinators and other users with interpreting the information on the tables. The 'Summary Tables' are broken down by 'City' and 'Department' tables, which have unique summary information relating to overall open data progress. The 'Critical Records Tables' provide definitions for headings relating to the Critical Records, which help inform data coordinators about the health of the datasets in their data inventories.

Conclusion

Throughout the design process, our group strived to create a process and tools that would promote sustained effort in updating data inventories. The scope of the product from the prototype to the final release has not changed, however, some additions and adjustments have been made to meet the needs and expectations of our partner. All documentation is now accessible in one handbook with a section for data coordinators that can be printed to PDF and distributed to them by the Data Services Team. The names of city departments were added to the Tracking and Summary Tables, and formulas were adjusted to account for the additional rows. Data validation for departments was incorporated into the Tracking and Summary Tables and additional data validation was added to the data inventory. The update overdue, due, and due soon calculations in the Critical Records Program were adjusted to be more understandable from a human perspective and the new functionality was documented in the respective user guide.

Challenges that our partner may encounter as they implement the Data Governance Health Status Checks include resistance by data coordinators to engage with the process and problems with the technical components of the product, the Tracking and Summary Tables and Critical Records Program, when real data is used or when changes in the process require changes to the formulas or Visual Basic code. To mitigate data coordinator resistance, we recommend setting reasonable goals for the number of datasets a department must identify and document between Data Governance Health Status Checks meetings. To resolve problems that arise with the *Tracking and Summary Tables*, the Data Services Team should involve people who have familiarity with Excel tables and formulas. Our understanding is that the team has sufficient access to people who know Excel. To resolve problems with or change functionality of the *Critical Records Program*, the Data Services Team should involve people who are familiar with Visual Basic for Applications (VBA) or who are comfortable with programming in general. Our understanding is that there is at least one person on the team who is familiar with VBA. The user guides that accompany these technical components will be helpful in interpreting their outputs as well as in understanding their purpose and functionality. The *User Guide: Critical Records Program* provides detailed explanations of different features of the program that will help anyone tasked with modifying the code. The VBA macro is also well documented.

Our partner is pleased with our work and has already started using the *Tracking and Summary Tables*. They mentioned that the *Data Coordinator Procedures* will be an important document as data coordinators work toward compliance with the Open Data Ordinance. They also answered the question posed at the beginning of this project, "Where are the dog licenses?" Dog licenses are managed by the Finance Department. The idea behind this question was that without good data management practices and open data, the city and stakeholders do not know what data they have, where data are located, or who manages the data. We trust that the process and tools we developed will serve as a foundation for the city to achieve their goal of consistent updating of data inventories that will lead to knowing where all the data are. As they move forward and become informed by their experiences, they can build on what we started to improve the process and tools to meet their needs.