Open Data Risk Analysis

|  |  |
| --- | --- |
| Open Data Candidate Information | |
| Data Candidate: | *Fill in the name of the data set under consideration* |
| Data Description: | *Brief description of the data set.* |
| Business Owner: | *Name of the division director who is responsible for the business side of the data set.* |
| Data Owner: | *Name of the division director who is responsible for the database and technical side of the data set.* |
| Technology: | *Name of database the data currently resides in* |
| Potential Audience: | City of Seattle Constituents |
| Current Audience: | *Who is currently using this data inside the City?* |

Contents

[Background 3](#_Toc286733568)

[Process for Evaluating Datasets 3](#_Toc286733569)

[Dataset Evaluation Process Diagram 4](#_Toc286733570)

[Roles and Responsibilities 5](#_Toc286733571)

[Dataset Teams: 6](#_Toc286733572)

[Guiding Principles Requirements 7](#_Toc286733573)

[Risk Factor 1: Data Security and Completeness 7](#_Toc286733574)

[Risk Factor 2: Data Complexity and Primacy 8](#_Toc286733575)

[Risk Factor 3: Data Maintenance 9](#_Toc286733576)

[Risk Factor 4: Data Usability and License 10](#_Toc286733577)

[Risk Factor 5: Metadata Completeness 11](#_Toc286733578)

[Risk Assessment 12](#_Toc286733579)

[Potential Benefits for Publishing Dataset 12](#_Toc286733580)

[Go/No Go Decision 12](#_Toc286733581)

[Acceptance: 12](#_Toc286733582)

[Data Publishing Agreement: 13](#_Toc286733583)

[Document Revisions 13](#_Toc286733584)

[Appendix A: Data Field Elements and Recommendation 13](#_Toc286733585)

[Appendix B: Table Purposes 13](#_Toc286733586)

[Appendix D: MetaData Form 13](#_Toc286733587)

# Background

Every City department owns data subject to FOIA requests. Some datasets can be easily identified and published, some require more careful scrutiny, and some datasets, after a careful examination has be performed, should not be published. The intent of this template is to help define and streamline publishing of data on data.seattle.gov. It is a guide for determining what to publish as open data and how to insure that the data is properly examined for security/privacy issues and to determine whether the dataset should be published. In the process, the data may be refined or modified. Examples of modifications might include renaming data column titles for clarity and understanding, deleting vestige data columns that no longer serve a purpose, or removing references to personal identification. The operative concept is common sense. A dataset that does not disclose personal indentifying information, sensitive/critical infrastructure, or competitively sensitive information is a candidate for publishing.

*<Enter a paragraph describing the data set candidate.>*

OPTIONAL: <*Enter a paragraph describing the data set team and the process the team went through to provide a recommendation to the Steering Committee. Also include the number of hours the process took.>*

# Process for Evaluating Datasets

Using this template should be an exercise in common sense. It is recommended that this process be used that follows the standard software lifecycle development process consisting of planning, analysis, design and maintenance. By following a modified SDLC approach the project team is ensuring that a complete analysis of a dataset is performed, a go/no go recommendation is made to a governing body and development and testing of a dataset occurs. The analysis stage is the emphasis for this deliverable and the design and maintenance stage will be completed once the dataset is approved for publication. Since this is a newly developed process there will be edits and revisions to the process as additional datasets are analyzed and go through the process.

The following diagram outlines the process for evaluating a dataset for publication:

# 

## Dataset Evaluation Process Diagram

# 

# 

# Roles and Responsibilities

The following roles and responsibilities should be included in every data set evaluation project. **Team members may have multiple roles and responsibilities.**

|  |  |
| --- | --- |
| Role | Responsibility |
| Project Manager  (or Dataset Coordinator) | Responsible for leading the dataset group through the dataset evaluation process. Works with the steering committee to inform them of the status, risks and/or issues during the data set evaluation process. Responsible for developing the final recommendation deliverable and risk analysis. |
| Business Owner | Sits on the steering committee. Responsible for the decision to publish or not publish data set based on dataset group’s recommendations. |
| Data Owner | Sits on the steering committee. Responsible for the decision to publish or not publish data set based on dataset group’s recommendations. |
| Data Expert | The technical representative for the dataset. Knows how the data is derived and formatted. Provides the tables and fields to the dataset group. Provides expertise for data extraction to data.seattle.gov |
| Business Expert | The business representative for the dataset. Knows the definitions for the tables and fields and provides the business knowledge around the usage of the tables and fields. |
| Business Analysts | The process and policy representative for the team. Provides the policy expertise to the team. |
| Public Information Officer | Provides the subject matter expertise for public information and is responsible for understanding the privacy risks and bringing them to the project manager for discussing with the steering committee. |
| DoIT Data.Seattle.Gov representative | Provides the dataset publishing expertise. |
| Customer Service representative | Provides the subject matter expertise for customer service and is responsible for understanding the customer service issues and bringing them to the core team and project manager for discussion with the steering committee. |

# Dataset Teams:

The following table outlines the recommended teams needed for evaluating datasets.

|  |  |
| --- | --- |
| Team | Responsibility |
| Steering Committee | This is the governing body who will evaluate the core team’s recommendation for publishing the data. This committee/role makes initial decision to publish based on a low risk analysis. The Steering Committee should consist of the technology and data owners. If risk analsys level evaluates to medium or high, the dataset publication decision should be elevated to the next management level. |
| Core Team | This is the core working group for the dataset. This team walks through the dataset evaluation process and conducts an analysis on each table and field being considered for publication. The core team is responsible for drafting the Open Data Candidate Requirements and Risk Evaluation deliverable as well as the corresponding Data Field Elements and Recommendation and Table Purpose deliverables. |
| Decision Maker(s) | This is the team (or person) responsible for making the final decision to publish a dataset to data.seattle.gov. This team is determined by the Risk Analysis Profile. |

# Guiding Principles Requirements

*Fill out each requirement for per principle.*

## Risk Factor 1: Data Security and Completeness

|  |  |  |
| --- | --- | --- |
| # | **Details** | |
| 1. | **Definitions:**  **Data Security:**  The following types of information are prohibited from disclosure through the open data initiative:   * Personally identifiable information * Personal health information * Cardholder or other personal financial information * Data that are regulated and prohibited from unauthorized disclosure, e.g. justice information under CJIS * Critical infrastructure information, e.g. the mapped location of networking fiber * Any information that is exempt from public disclosure   **Data Completeness:**   * Within the confines of data security, the data set represents the complete set of fields and records | |
| 2. | **Requirements** | |
| 2.1 | List all data fields in original dataset  Enumerate the fields that need to be filtered out for publication if any.  General questions to ask for data security and completeness   * Does the dataset contain and data violating the City’s data security prohibition? * If so, can the dataset be filtered to remove the prohibited data? * Does dataset with prohibited data removed retain usefulness? * Does the dataset contain any legacy, obsolete data fields, or internal flags? | |
| 2.2 | **Exceptions:**  *Enter any exceptions to this requirement.* | |
| 3 | **Issues/Action Items** | |
| 3.1 | *Enter any issues or action items that need to be conducted prior to the finalization of the dataset.* | |
| 4. | **Existing Policies/Guidelines/Standards** | |
|  | Yes \_\_\_\_  *Note if there are any existing policies/guidelines/standard.* | No \_\_\_\_  *Note if any policies/guidelines/standard need to be developed prior to publishing the dataset.* |
| 5. | **Risk Scale – Value to be placed on Risk Analysis Visio Diagram** | |
|  | Risk = *Enter the applicable risk*  **Low** = Data set to be published as is.  **Medium** = Some work needs to be done prior to data set being published. Some fields are recommended for exclusion.  **High** = All data fields contain some element of risk or too much data clean up needed prior to publishing data. | **Risk Rational:**  *Enter the rational for the risk decision.* |

Check when completed

## Risk Factor 2: Data Complexity and Primacy

Check when done

|  |  |  |
| --- | --- | --- |
| # | **Details** | |
| 1. | **Definition:**  Data is as collected at the source, with the highest possible level of granularity, not in aggregate or modified forms. | |
| 2. | **Requirements** | |
| 2.1 | Data should be at the detail level and if it is not at the detail level there should be some explanation as to why it is not available at the detail level. This principle also implies that no data fields should be combined together to make a new field and the same for tables. The idea is to show the data in the rawest format and let public developers do the mashing of data. However, in the absence of primary data or primary data with security constraints, aggregated data is preferable to no data.  General questions to ask for data complexity:  Is data at the detail level?  Are there dependencies on other related tables?  If summary data is available, can the source data be published with the summary data? | |
| 2.2 | **Exceptions:**  *Enter any exceptions to this requirement.* | |
| 3 | **Issues/Action Items** | |
| 3.1 | *Enter any issues or action items that need to be conducted prior to the finalization of the dataset.* | |
| 4. | **Existing Policies/Guidelines/Standards** | |
|  | Yes \_\_\_\_  *Note if there are any existing policies/guidelines/standard.* | No \_\_\_\_  *Note if any policies/guidelines/standard need to be developed prior to publishing the dataset.* |
| 5. | **Risk Scale – Value to be placed on Risk Analysis Visio Diagram** | |
|  | Risk = *Enter the applicable risk*  **Low** – data is granular, not in aggregate or modified forms  **Medium** – some data is in summary and would have to be broken down into detail information. Or, only summary data is available.  **High –** Data is highly complex and/or in a relational form of multiple tables and requires extensive effort to publish as a single entity | Risk Rational:  *Enter the rational for the risk decision.* |

Check when completed

## 

## Risk Factor 3: Data Maintenance

|  |  |  |  |
| --- | --- | --- | --- |
| # | **Details** | | |
| 1. | **Definition:**  Data is made available as quickly as necessary to preserve the value of the data. | | |
| 2. | **Requirements** | | |
| 2.1 | If the dataset is consistently updated then the dataset extract should be updated on the same timeline. There should be no significant lags in providing new data to constituents.  General questions to ask for the availability of data in a timely manner:   * Are transactions occurring daily * What do the constituents want? * Is there an impact to the process or the production cycle? * If data extraction is identified as difficult, are there alternative methodologies to simplify data extraction? | | |
| 2.2 | **Exceptions:**  *Enter any exceptions to this requirement.* | | |
| 3 | **Issues/Action Items** | | |
| 3.1 | *Enter any issues or action items that need to be conducted prior to the finalization of the dataset.* | | |
| 4. | **Existing Policies/Guidelines/Standards** | | |
|  | Yes \_\_\_\_  *Note if there are any existing policies/guidelines/standard.* | | No \_\_\_\_  *Note if any policies/guidelines/standard need to be developed prior to publishing the dataset.* |
| 5. | **Risk Scale – Value to be placed on Risk Analysis Visio Diagram** | | |
|  | Risk = *Enter the applicable risk*  **Low**  - data is easy to extract, data can easily be made available in a timely fashion  **Medium** – data is more difficult to extract and a scheduled data extract is needed. No or minimal impact to production cycles for application.  **High** – data will be difficult to extract and impacts production cycle for application. | Risk Rational:  *Enter the rational for the risk decision.* | |

Check when completed

## 

## Risk Factor 4: Data Usability and License

|  |  |  |  |
| --- | --- | --- | --- |
| # | **Details** | | |
| 1. | **Definition:**  Data is available to the widest range of users for the widest range of purposes  Data is reasonably structured to allow for automated processing.  Data is available to anyone, with no requirement of registration and data is available in a format over which no entity has exclusive control.  Data is also not subject to any copyright, patent, trademark or trade secret regulation. Reasonable privacy, security and privilege restriction may be allowed. | | |
| 2. | **Requirements** | | |
| 2.1 | There should not be any barriers on constituent’s ability to access the data. Barrier’s could be the technology, size of the dataset.  General questions to ask for data usability and license   * Are there restrictions to use based on dataset size? * Are there no preexisting rules, regulations or policies prohibiting the data to be published? * Is dataset available in a machine readable form? * Is the dataset in a proprietary file format? * Can the dataset be released to public domain, free of any license? | | |
| 2.2 | **Exceptions:**  *Enter any exceptions to this requirement.* | | |
| 3 | **Issues/Action Items** | | |
| 3.1 | *Enter any issues or action items that need to be conducted prior to the finalization of the dataset.* | | |
| 4. | **Existing Policies/Guidelines/Standards** | | |
|  | Yes \_\_\_\_  *Note if there are any existing policies/guidelines/standard.* | | No \_\_\_\_  *Note if any policies/guidelines/standard need to be developed prior to publishing the dataset.* |
| 5. | **Risk Scale – Value to be placed on Risk Analysis Visio Diagram** | | |
|  | Risk = *Enter the applicable risk*  **Low** – No barriers to data set and low risk to data set usage by public  **Medium** – No barriers to data set and some risk to data set usage by the public  **High** – Barriers to data and high risk to data set usage | Risk Rational:  *Enter the rational for the risk decision.* | |

Check when completed

## 

## Risk Factor 5: Metadata Completeness

|  |  |
| --- | --- |
| # | **Details** |
| 1. | **Definition:**  Metadata or data describing the dataset must be attached. |
| 2. | **Requirements**  A contact person must be designated to respond to constituents trying to use the data.  A contact person must be designated to respond to constituent complaints about violations of privacy or violations of the principles  All dataset columns must be documented in plain English for purpose and datatype  Spatial datasets must contain the spatial extent  Spatial datasets must contain the projection definition  All technical issues or questions will be handled by Socrata. |
| 2.1 | **Exceptions:**  **Enter any exceptions to this requirement.** |
| 3.0 | **Issues/Action Items** |
| 3.1 | *Enter any issues or action items that need to be conducted prior to the finalization of the dataset.* |
| 5. | **Risk Scale – Value to be placed on Risk Analysis Visio Diagram** |
|  | Risk = *Enter the applicable risk*  **Low** – Data set is fully documented and will require minimum customer service care  **Medium** – Data set is difficult to understand but can be handled internally if issues arise  **High** – Data set being published will require special care when published or metadata is not available. |

Check when completed

# 

# Risk Assessment

*Include a brief description of the risk associated to this data set and the final risk profile chart.*

# Potential Benefits for Publishing Dataset

*Describe any potential benefits that could be realized with the publication of the dataset. This could include additional financial benefits to the City or process benefits.*

# Go/No Go Decision

*Data team recommendation to the Steering Committee for moving forward or not moving forward with publishing data set on data.seattle.gov*

# Acceptance:

We, the undersigned decision makers, have reviewed this document and approve of the Go/No Go Dataset Publishing Recommendations and the deliverable:

|  |  |
| --- | --- |
| **Executive Sponsors:** | |
|  |  |
| Signature: | Date: |
|  |  |
| Signature: | Date: |
|  |  |
| Signature: | Date: |
|  |  |
| Signature: | Date: |

# 

# Data Publishing Agreement:

I, the undersigned decision maker, have reviewed this document and the risk analysis and approve of the Go/No Go Dataset publishing recommendation and agree to publish the data on data.seattle.gov:

|  |  |
| --- | --- |
| **Executive Decision Maker:** | |
|  |  |
| Signature: | Date: |

# Document Revisions

|  |  |  |
| --- | --- | --- |
| Version # | Revised Date | Description |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# [Appendix A: Data Field Elements and Recommendation](file:///\\Dea01_v2_server\v2\Data\Summit\Pm\Project%20Management\Projects\2010_In%20Progress\I17_FAS.DATA.SEATTLE_Execution\04_Executing%20&%20Controlling\deliverables\AppendixA_datafieldelementsandrecommendations.docx)

*Shows all the tables and fields associated to the dataset and provides decision makers a recommendation of whether or not fields should be included in the published data set.*

[Appendix B: Table Purposes](file:///\\Dea01_v2_server\v2\Data\Summit\Pm\Project%20Management\Projects\2010_In%20Progress\I17_FAS.DATA.SEATTLE_Execution\04_Executing%20&%20Controlling\deliverables\Appendix%20B%20_FAS%20SLIM%20Table%20Purposes.docx)   
*Describes the purpose of each table used for publishing in the raw data set.*

**Appendix C: Data Publishing and Maintenance Plan**

*Include if there is a go decision. Show a high level schedule for publishing the data set, define who the data set contacts are, and decided publishing timeline (weekly, monthly, etc).*

# [Appendix D: Metadata Form](file:///\\Dea01_v2_server\v2\Data\Summit\Pm\Project%20Management\Projects\2010_In%20Progress\I17_FAS.DATA.SEATTLE_Execution\04_Executing%20&%20Controlling\deliverables\AppendixD_metadata_businesslicensedata.docx)