Open Data Readiness Toolkit

Evaluating and Inventorying Departmental Data Sets

# Why is Evaluating and Inventorying Data Sets important?

Open data programs build on established digital data sources and information management procedures within government where they already exist. Where data is only available in paper form it will be hard to release as open, reusable data quickly and cheaply. Conversely, good existing information management practices within the jurisdiction can make it much easier to find data and associate metadata and documentation, identify business ownership, assess what needs to be done to release it as open data. It also makes it easier to put processes in place that make the release of data a sustainable, business-as-usual, downstream process as part of day-to-day information management.

# What does good look like?

All data would be held digitally. The jurisdiction would have a comprehensive inventory of its data assets and each dataset would have comprehensive metadata and supporting documentation. All data assets should be entered into an Information Asset Registry (IAR: Open Data IAR Departmental Template).

## Questions to Ask

Are there special policies/laws on the management of this department’s information assets?

* Does the department have a governance process for information and data security?
* What policies or standards exist on the provenance, accuracy and quality control of data?
* What is the retention schedule for the information assets of this department?
* Does the department have a coherent view of its data assets?
* Is there an inventory of data assets?
* If there is no government inventory, do any key individuals have one?
* How comprehensive and up-to-date is the inventory(s)?
* Are there established metadata standards?
* Do accurate metadata records describe the data assets?

## How and where is department data held?

* How much current data is in digital and reuse forms (rather than paper, microfilm, image etc. forms)? What about legacy data (including both historical records and past transactions)?
* To what technological formats and standards is the department’s data held? Are they open or proprietary formats? Is there an interoperability framework?
* If public service information is held by contractors or outsourced suppliers, does the department responsible have rights to access and distribute it? Do procurement laws/regulations retain ownership of such information with the government?
* How is data archived once it has ceased to be used operationally? Who has authority/ability to access archived data?

## What data is already published and on what conditions?

* What datasets are published? In what format, and is it re-usable?
* Is information available in fixed files, or also through APIs?
* What income (if any) is generated? What are the costs of administrating any charges?
* What conditions are imposed? For what reason?
* Is data on department webpages downloadable?
* Is this a designated department with established capabilities in data management (The Budget Office or Information Technology)? This could give leadership to a wider initiative.

## Useful existing departmental capabilities would include:

* Data collection, curation, management and publication
* Use of well-documented, standard workflow for data management
* Data analytics
* Privacy safeguarding and anonymization

# Assessment of Department Data and Recommended Actions

|  |  |  |  |
| --- | --- | --- | --- |
| **Question Area** | **Importance** | **Assessment** | |
| Red/Amber/  Green rating\* | Commentary |
| Policies on Information Management | Medium |  |  |
| View of Data Holdings | High |  |  |
| How is data held | Very High |  |  |
| Existing publication of data | Medium |  |  |
| Agencies with existing capability | High |  |  |
| Availability of key datasets | Very High |  |  |
|  |  |  |  |
| OVERALL | High |  |  |

# Data Recommendations for Action Plan

|  |  |  |  |
| --- | --- | --- | --- |
| Action | Responsibility | Timescale | Comments |
| 1. |  |  |  |
| 2. |  |  |  |
| 3. |  |  |  |

# Availability of Key Datasets

Consider the following areas. Which department is responsible for each, and what relevant skills, capacity and experience does each department have? How difficult would it be to release this data in reusable form? What would be the key obstacles? ***(Complete “key dataset template” for as many of the following as possible, and summarize results in “key dataset findings and recommendations”)***

1. Budget data (both at OMB and individual department level)
2. Disaggregated expenditure and grant data
3. Statistical data from individual departments publishing recognized KPI statistics
4. Census data
5. Governance data including records of proceedings, drafts under debate and enacted version of legislation
6. Procurement data (who was awarded what) and contract data (the documents and details of the deal)
7. Data on public facilities including schools, hospitals, police stations, public toilets, bus stops, libraries, government offices etc. - location and services available
8. Public service delivery and performance data at the level of individual facility, hospital/clinic etc.
9. Transport data including roads and public transport
10. Crime data to the level of individual crimes and their locations
11. Reports of inspections and official decisions and rulings in reusable form
12. Official registers - company, charities, cadastral/land ownership etc.
13. Geospatial information - maps, address registers, points of interest.
14. Weather data
15. Construction data (permits, zoning)
16. Real estate data (sales, listings, taxes, other property-specific data)

# KEY DATASET TEMPLATE

|  |  |
| --- | --- |
| **Dataset title and description** |  |
| **Dept/Agency which stewards the dataset** |  |
| ***State of the Data (questions to be submitted to data owner in advance of assessment)*** | |
| Does the data exist? Is it in digital and reusable form? |  |
| What format(s) is the data currently in? |  |
| What is the “granularity” of the data? How aggregated or de-aggregated is it? |  |
| Is data archived in its raw, more granular form? |  |
| How often is the data updated? |  |
| Is the data available as a complete set (raw data download) or only as the result of individual queries? |  |
| How is the production and distribution of data currently funded? |  |
| To what extent is the data available to the public, to businesses or to others outside government? |  |
| If the data is available to others, is it free or charged-for? |  |
| If the data is available to others, what restrictions are placed on their use, and why? |  |
| What relevant skills, capacity and experience does the owning agency have? |  |

|  |  |  |
| --- | --- | --- |
| ***Feasibility assessment of Release (assessment to be done as part of the Readiness Assessment, prioritized as time allows)*** | | |
| Policy feasibility: What are the key policy obstacles, which would have to be overcome? | High/Medium/Low | Comment |
| Technical feasibility: What are the key obstacles, which would have to be overcome? | High/Medium/Low | Comment |
| Institutional feasibility: What would be the key obstacles in terms of capacity and capability? | High/Medium/Low | Comment |

# KEY DATASET FINDINGS AND RECOMMENDATIONS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Dataset** | **Feasibility** | | | **Recommendation\*** |
| **Policy** | **Technical** | **Institutional** |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

## Criteria

1. Easy/Quick Win
2. Priority/Medium Term
3. Difficult/Longer Term