Table of Contents

Background .............................................................................................................................................. 1
Requirements ........................................................................................................................................... 1
Enterprise Data Inventory ..................................................................................................................... 1
Public Data Listing ................................................................................................................................. 2
Customer Engagement ........................................................................................................................... 3
Non-Releasable Data ............................................................................................................................... 3
Roles and Responsibilities ....................................................................................................................... 4
Concept of Operation .............................................................................................................................. 4
Schedule ................................................................................................................................................... 5
OPEN DATA PLAN

U.S. DEPARTMENT OF STATE

Background

OMB Memorandum M-13-13 Open Data Policy-Managing Information as an Asset, published on May 9, 2013, establishes a framework to help institutionalize the principles of effective information management at each stage of the information's life cycle to promote interoperability and openness. The Open Data Policy has as its goals to increase operational efficiencies at reduced costs, improve services and support mission needs, to safeguard personal information and to increase public access to valuable government information. For data to be open, it must be machine readable using open formats, follow open data standards, use open licenses, and adhere to a government-wide common core metadata standard.

Requirements

The Open Data plan describes how the Department of State will meet the following five initial requirements of M-13-13, which are due November 30, 2013:

- Create and maintain an Enterprise Data Inventory (EDI)
- Create and maintain a Public Data Listing
- Create a process to engage with customers to help facilitate and prioritize data release
- Document if data cannot be released
- Clarify roles and responsibilities for promoting efficient and effective data release

Enterprise Data Inventory

The Department currently manages their inventory of agency information resources through the iMatrix system. iMatrix is the single authoritative source and system of record for Department systems (applications, networks and websites). There are entries currently for approximately 360 Department systems in iMatrix. It is the source for responding to a number of external reporting requirements, enabling the Department to construct vital portions of the Enterprise Architecture, and supports the Cyber Security Program, including the systems authorization (Certification and Accreditation) process. It also supports the Department's eGov initiatives and is helping streamline business processes.

To fulfill the requirement for an inventory of all enterprise data, the iMatrix system will be enhanced to include space in the system record for an inventory of its datasets. This will be accomplished by defining a new asset type called DATASET. Once this is implemented, iMatrix will become the Department of State’s Enterprise Data Inventory (EDI) and system owners will be able to enter information on the data assets they manage. The DATASET asset type that will be added to the data structure of the iMatrix is shown in Figure 1.

The datasets for the existing systems will be populated in the second quarter of FY 2014 through a Department-wide data call for system owners to update their entries in iMatrix. System owners will be required to enter the dataset information on systems created after the EDI implementation date as part of their initial iMatrix system entry. Data Stewards will also be approached through the Application and Data Coordination Working Group (ADCWG). The ADCWG is comprised of a broad array of stakeholder bureaus from across the Department, and is working to standardize data so that information
systems can communicate more effectively through central data tables and hardware, reducing the need for ad hoc data calls. All new system datasets will be routed through the ADCWG to ensure adherence to data quality standards and will be entered into the EDI at that time.

<table>
<thead>
<tr>
<th>OMB Required Fields*</th>
<th>DoS Specific Fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Size (in MB) – to determine storage</td>
</tr>
<tr>
<td>Description</td>
<td>Growth % - to determine future needs</td>
</tr>
<tr>
<td>Tags (pull-down list)</td>
<td>FISMA Categorization</td>
</tr>
<tr>
<td>Last Update</td>
<td>PII Information</td>
</tr>
<tr>
<td>Publisher</td>
<td>Mobile enabled/ready</td>
</tr>
<tr>
<td>Contact Name</td>
<td>Information Area (pull-down)</td>
</tr>
<tr>
<td>Contact Email</td>
<td>Information Group (pull-down)</td>
</tr>
<tr>
<td>Unique Identifier (assigned by iMatrix)</td>
<td>Note: Additional data elements can be captured so that DoS can obtain valuable insights into information asset</td>
</tr>
</tbody>
</table>

Data entered into the EDI will also adhere to the metadata standards set up in the Enterprise Metadata Repository. The Enterprise Metadata Repository will store additional metadata information like record layout, column types, permissible values and usage to support the standardization of data across the Department. If the metadata to be used in the EDI is not already in the Enterprise Metadata Repository, a registry record for the new data type will be created. This will standardize the format and use of metadata in the EDI.

**Public Data Listing**

The Department will publish a Public Data Listing containing all data assets that are, or could be made available to the public. This Public Data Listing will be a subset of the Department’s EDI and will

---

* Metadata is based on: Project Open Data (http://project-open-data.github.io/schema/)
** Determines if additional data fields are required
+ Additional data fields used if data is to be released to the Public

---

*Figure 1 – Dataset Asset Type in iMatrix*
allow the public to view the open data assets and track the progress made as additional data assets are published. To make the public aware of data that is not releasable and the process by which these data may be obtained, entries in the Public Data Listing may include the metadata on data that is not releasable, but not the actual data.

The Public Data Listing will be used to dynamically populate Data.gov which allows the public to use a single search engine to find data assets generated and held by the U.S. Government. Data.gov will automatically aggregate the agency-managed Public Data Listings into one centralized location, using the common core metadata standards and tagging to improve the user ability to find and use government data. The Public Data listing will be located on the www.State.gov/data page and be contained in a single JSON file. The Public Data listing will be refreshed quarterly at a minimum.

**Customer Engagement**

Identifying and engaging with key data customers to help determine the value of federal data assets can help agencies prioritize those of highest value for quickest release. Customers will be engaged through blog entries, email, forms on the www.State.gov/open web page, and other means as appropriate. Customers include public as well as government stakeholders. Internal customers will use blogs, email and Corridor (the Department social media site) to interact with data owners directly. The Department will evaluate public and private input and reflect on how to incorporate it into their data management practices. The Department will regularly review its evolving customer feedback and public engagement strategy and develop criteria for prioritizing the opening of data assets, accounting for factors such as the quantity and quality of user demand, internal management priorities, and agency mission relevance.

**Non-Releasable Data**

The Open Data Policy requires agencies to strengthen and develop policies and processes to ensure that only the appropriate data are publicly available. If the Department determines the data should not be made publicly available because of law, regulation, or policy or because the data are subject to privacy, confidentiality, security, trade secret, contractual, or other valid restrictions to release, it must document the determination in consultation with the Office of the Legal Advisor (L Bureau). Datasets will belong to one of three categories: public, restricted public, and non-public. The descriptions of these categories are the following:

- **Public**: Data asset is or could be made publicly available to all without restrictions.

- **Restricted Public**: Data asset is available under certain use restrictions. The accessLevelComment field in the metadata must be filled in with details on how one can obtain access.

- **Non-Public**: Data asset is not available to members of the public. This category includes data assets that are only available for internal use by the Federal Government, such as by a single program, single agency, or across multiple agencies. The accessLevelComment field in the metadata must contain an explanation for the reasoning behind why these data cannot be made public for non-public datasets.
Roles and Responsibilities

The roles and responsibilities are listed for the following Open Data participants:

- **System Owners** – The System Owner has overall responsibility for all aspects of the information system that holds data. The registered System Owner is identified in iMatrix. The System Owner is responsible for entering all of the descriptive metadata on the system including the datasets created and maintained by the information system.

- **Data Stewards** – The Data Steward is the person that is responsible for the data entered into the information system and ensures that the data entered is correct and meets quality requirements for currency and accuracy. The Data Steward makes the decision as to whether the data should be Public, Restricted Public, or Non-Public. The Data Steward prepares any documentation required to establish a dataset as Restricted Public or Non-Public.

- **iMatrix System Owner (IRM)** – The iMatrix system owner maintains the iMatrix system which contains, as one of its functions, the Enterprise Data Inventory.

- **E-Government Program Board** – Ensures that IT proposals meet Department’s and OMB’s IT and E-Gov strategic principles, which includes the Open Data policy.

- **ITCCB** – The Information Technology (IT) Change Control Board (CCB) manages changes to the Department of State’s global IT environment. As such, the ITCCB is responsible for ensuring that new IT systems and changes to existing IT systems adhere to the Open Data policy.

- **Application and Data Coordination Working Group (ADCWG)** – The ADCWG has an Enterprise Data Quality Initiative that addresses the accessibility, reusability, reliability and overall quality of enterprise data. The metadata entered into the EDI and the data entered into the datasets will have to follow directives associated with this initiative.

- **Chief Information Officer** – The CIO is ultimately responsible for the department-wide implementation of all Open Data requirements.

Concept of Operation

The System Owner (new or existing system) will identify all key data sets that can be created and published. The System Owner captures the core metadata information about the data set in iMatrix. The extended metadata, like record layout or permissible values, are entered into the Enterprise Metadata Registry. When entering the metadata the System Owner consults with Data Steward about the correct categorization of the data: public, restricted public, or non-public. Legal will have the responsibility to make the final determination if the data can be open. The iMatrix system owner will designate a user that will perform the metadata extraction process on the EDI, and subsequently process the data into a JSON file. The JSON file will be published on the [www.state.gov/data](http://www.state.gov/data) page. This process will be done periodically, and not less than quarterly at the start.

The concept of operations is shown in Figure 2.
Figure 2 – Concept of Operations

Schedule

The Department will start with the datasets owned by the organizations shown in Table 1.

<table>
<thead>
<tr>
<th>Owner</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMART</td>
<td>Contains various information on data tagging and the policies being transmitted</td>
</tr>
<tr>
<td>ILMS</td>
<td>Contains various information that is used for assisting bureaus and offices in better managing the procurement</td>
</tr>
<tr>
<td>MRD</td>
<td>Contains some of the master reference datasets that are published for all systems within State to use</td>
</tr>
<tr>
<td>SPD</td>
<td>Has all of the information that has already been published through data.gov</td>
</tr>
<tr>
<td>PA</td>
<td>Contains the different reports and information that is published through <a href="http://www.state.gov">www.state.gov</a></td>
</tr>
<tr>
<td>DRL</td>
<td>Owner of reports and data related to Human Rights</td>
</tr>
<tr>
<td>INL</td>
<td>Contains reports and data that has been published through their website</td>
</tr>
</tbody>
</table>

Table 1 – Bureaus or Offices to be entered into the EDI

Every quarter the Department will target specific bureaus/offices and IT systems within its portfolio to reach out and communicate the Open Data Policy and obtain the datasets that they are currently producing. The list of the datasets will be made available through the Enterprise Data Inventory. Once
it is initially entered – the dataset owner will be responsible for the update and maintenance of the dataset and the associated metadata.

The schedule for the implementation of Open Data is shown in Table 2.

<table>
<thead>
<tr>
<th>Milestone</th>
<th>Description</th>
</tr>
</thead>
</table>
| 1         | • Title: Initial Delivery  
            • Description: The initial delivery of the Open Data Plan, the Schedule, the Enterprise Data Inventory and the Public Data Listing  
            • Date: November 30, 2013  
            • Number of datasets: 113  
            • Open Datasets: 99 |
| 2         | • Title: 1st Quarterly Update  
            • Description: Update Open Data Plan, Schedule, Enterprise Data Inventory and Public Data Listing  
            • Date: February 28, 2014  
            • Datasets Expanded: 36 (149 total datasets)  
            • Datasets Enriched: 18  
            • Datasets Open: 9 (108 total open datasets) |
| 3         | • Title: 2nd Quarterly Update  
            • Description: Update Open Data Plan, Schedule, Enterprise Data Inventory and Public Data Listing  
            • Date: May 31, 2014  
            • Datasets Expanded: 72 (221 total datasets)  
            • Datasets Enriched: 18  
            • Datasets Open: 9 (117 total open datasets) |
| 4         | • Title: 3rd Quarterly Update  
            • Description: Update Open Data Plan, Schedule, Enterprise Data Inventory and Public Data Listing  
            • Date: August 30, 2014  
            • Datasets Expanded: 72 (293 total datasets)  
            • Datasets Enriched: 36  
            • Datasets Open: 18 (126 total open datasets) |
| 5         | • Title: 4th Quarterly Update  
            • Description: Update Open Data Plan, Schedule, Enterprise Data Inventory and Public Data Listing  
            • Date: November 30, 2014  
            • Datasets Expanded: 72 (365 total datasets)  
            • Datasets Enriched: 36  
            • Datasets Open: 18 (144 total open datasets) |

Table 2 – Schedule

At the end of one year, at least 85% of the systems’ datasets will be entered into the EDI and at least 30% of the entered datasets will be made publicly available.