SBA

Open Data Program

SMALL BUSINESS ADMINISTRATION STANDARD OPERATING PROCEDURE National		

SUBJECT: SBA Implementation of OMB Memorandum M-13-13, *Open Data Policy-Managing Information as an Asset*.	S.O.P.		REV
	SECTION 90	NO. 81	

INTRODUCTION

Purpose: To outline the policies and procedures for the SBA Agency Offices

Personnel Concerned: All SBA Agency Offices

Directives Canceled: None

Originator: Office of Chief Information Officer

AUTHORIZED BY: Renee Macklin Chief Information Officer		EFFECTIVE DATE October 2, 2014
		PAGE 1

SBA Form 989 (5-90) Ref: SOP 00 23

Table of Contents

Open Data Program

I. Purpose

This standard operating procedure (SOP) serves two main purposes:

> A. It establishes the U.S. Small Business Administration's (SBA) Open Data Program, based on the requirement set forth by the enactment of the Office of Management and Budget (OMB) Memorandum M-13-13, *Open Data Policy, Managing Information as an Asset*.
>
> B. It defines the SBA roadmap for the implementation of the SBA Open Data policy, including detailed implementation procedures, roles and responsibilities of SBA offices involved in the implementation of the procedures, and definitions of key terms.

The SOP will assist the Agency offices in identifying and publishing data sets after determining what information may or may not be released. This determination will ensure that data submitted to OMB for publishing to Data.gov contains useful information about the activities and performance of SBA programs and operations. SBA must take into consideration data quality with regards to currency and usability, as well as privacy in terms of protecting information that cannot be released to the public.

The Open Data Working Group (ODWG) at SBA created this SOP that defines a process to implement and execute Executive Order 13642, "Making Open and Machine Readable the New Default for Government Information" and OMB M-13-13 "Open Data Policy – Managing Information as an Asset".

For definitions of terms used in this SOP, refer to Appendix A: Definitions

II. Authoritative References

- **Clinger-Cohen Act of 1996** – Requires the establishment of processes for executive agencies to analyze, track, and evaluate the risks and results of major investments in IT and requires reporting on the net program performance benefits achieved by agencies.
- **Public Law 106-554, Section 515** – "Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies".
- Executive Order 13642 of May 9, 2013, Making Open and Machine Readable the New Default for Government Information
- OMB Memorandum M-13-13, Open Data Policy – "Managing Information As An Asset" (May 9, 2013)
- OMB Memorandum M-09-12, "Memorandum on Transparency and Open Government" (Jan. 21, 2009)
- OMB Memorandum M-10-06, *Open Government Directive* (Dec. 8, 2009)
- OMB Memorandum M-06-02, Improving Public Access to and Dissemination of

Government Information and Using the Federal Enterprise Architecture Data Reference Model (Dec. 16, 2005)

- Office of Management and Budget, Digital Government: "Building a 21st Century Platform to Better Serve the American People" (May 23, 2012)
- OMB Memorandum M-10-23, Guidance for Agency Use of Third-Party Websites and Applications (June 25, 20I0)
- Executive Order I 3556, Controlled Unclassified Information

In addition to complying with the Privacy Act of 1974, the E-Government Act of 2002, the Federal Information Security Management Act (FISMA), and the Confidential Information Protection and Statistical Efficiency Act (CIPSEA), the SBA should implement information policies based on the Fair Information Practice Principles and NIST guidance on Security and Privacy Controls for Federal Information Systems and Organizations.

III. Scope

This SOP applies to all Agency offices that create, receive, and maintain data. It covers all IT acquisitions, investments, programs, projects, and operations.

IV. Policy

This Standard Operating Procedure (SOP) provides a roadmap and detailed implementation plan to assist the U.S. Small Business Administration (SBA) in implementing the objectives of the Office of Management and Budget (OMB) Memorandum M-13-13, *Open Data Policy-Managing Information as an Asset*.

Per OMB's M-13-13 memorandum, SBA is required to maintain an Enterprise-wide Data Inventory that accounts for all data created or collected. This includes, but is not limited to, data used in SBA information systems. SBA will continually improve the usefulness of the Inventory by expanding, enriching, and opening the Inventory to the public.

This SOP defines a process by which SBA data is approved and published and also provides guidance for maintaining the Enterprise-wide Data inventory.

The Open Data SOP provides process and procedures for:

- Agency Offices to identify data for dissemination to the general public, or determining the rationale for withholding certain data from the public.
- Data submitted to Data.gov that represents useful information about the activities and performance of SBA programs and operations, with consideration to issues of quality and privacy.
- Contributions to the enrichment of the description of data through improved metadata and by increasing the quantities of data shared with other Agencies and the public.
- Maintaining an Internal Enterprise Inventory (IEI) repository that accounts for data used in SBA information systems.

This SOP incorporates a set of structures and processes designed to manage the lifecycle of SBA structured and unstructured data. It provides the guidance necessary to ensure that:

- All new IT systems meet the requirements identified in OMB Memorandum M-13-13.
- Best practices are used to help facilitate Open Data deliverables in a timely fashion.
- Roles and responsibilities are outlined with respect to managing SBA data for quality and privacy, data submissions, dataset approvals and release; to address requests from the General Accountability Office (GAO) and Inspector General (IG); and to work with the OMB and comply with M-13-13 requirements.
- Agency-wide data with supporting logical, physical models and data-dictionaries are well-described.
- The IEI repository will enrich the quality of metadata by making datasets re-usable, manageable, and easily discoverable.

SBA must ensure that this underlying data is only released to the public after a full analysis of privacy, confidentiality, security, and other valid restrictions specified by laws and policies.

In accordance with this SOP and Federal contract requirements, the Program Managers must work with the Contracting Officers to ensure that procurement for services and products involving new and existing SBA Data Information Assets are in accordance with OMB Memorandum M-13-13, *Open Data Policy-Managing Information as an Asset* and Executive Order 13642 *Making Open and Machine Readable the New Default for Government Information*.

V. Responsibilities

The following chart details the roles and responsibilities involved in SBA's Open Data Initiative:

Roles	Responsibilities
OCIO Technical Team	- Provide support to Agency Offices through the Solution Architect Data Group (SADG). - Provide support throughout the process of preparing datasets for publishing. - Assist with the Data publication process. (For details about the responsibilities of the OCIO Technical Team, refer to **Section VI.AB OCIO Role (OCIO-SADG Swim Lane**)
Agency Offices	- Responsible for assigning a Data Steward to represent the Agency Office and their System Owners. - The custodians and authoritative sources of data at SBA via System owners and data stewards. - Responsible for initiating and/or responding to requests for data publication. - Consult with SBA stakeholders to approve data publication. - Responsible for the process of data publishing, metadata, and supporting document compilation with support from the OCIO team and Solution Architecture Data Group.

	(For details about the responsibilities of Agency Offices, refer to **Section VI.C Agency Office Roles (Agency Offices Swim Lane)**
Solution Architecture Data Group (SADG)	• Provide technical help to the various Agency Offices in addition to coordinating with the OCIO team and OMB, as necessary, to publish the datasets. • Primary contact for the OMB with respect to Open Data activities. (For the full list of SADG responsibilities, refer to **Section VI.D Solution Architect Data Group (SADG) Activities (Solution Architect Data Group Swim Lane)**
Compliance Reporting	• Review all SBA submissions of approved datasets, along with metadata and supporting documentation, according to their internal review process. • Publish the SBA data listing to Data.gov • Respond to compliance reporting requirements.

In addition to the above table, Task level roles and responsibilities are detailed in Appendix D: RACI Chart.

VI. Procedures

This SOP covers the procedures required to publish datasets (including data extracts, web services, and widgets) on the SBA.gov/data website and on the OMB Max Portal. The SBA maintains its listing of data in machine-readable format (such as "JSON") called the "Public Data Listing" (PDL) at http://www.SBA.gov/data.json. An Enterprise Data Inventory (EDI) of all SBA data in machine-readable format that are made publicly-available is also posted on the OMB Max Portal (https://max.omb.gov) which then feeds to Data.gov.

The following section describes the overall workflow for dataset submissions and approvals. Additionally, a Requirements Checklist and a Responsible, Accountable, Consulted and Informed (RACI) Chart has been developed and are available in Appendix B and C, respectively.
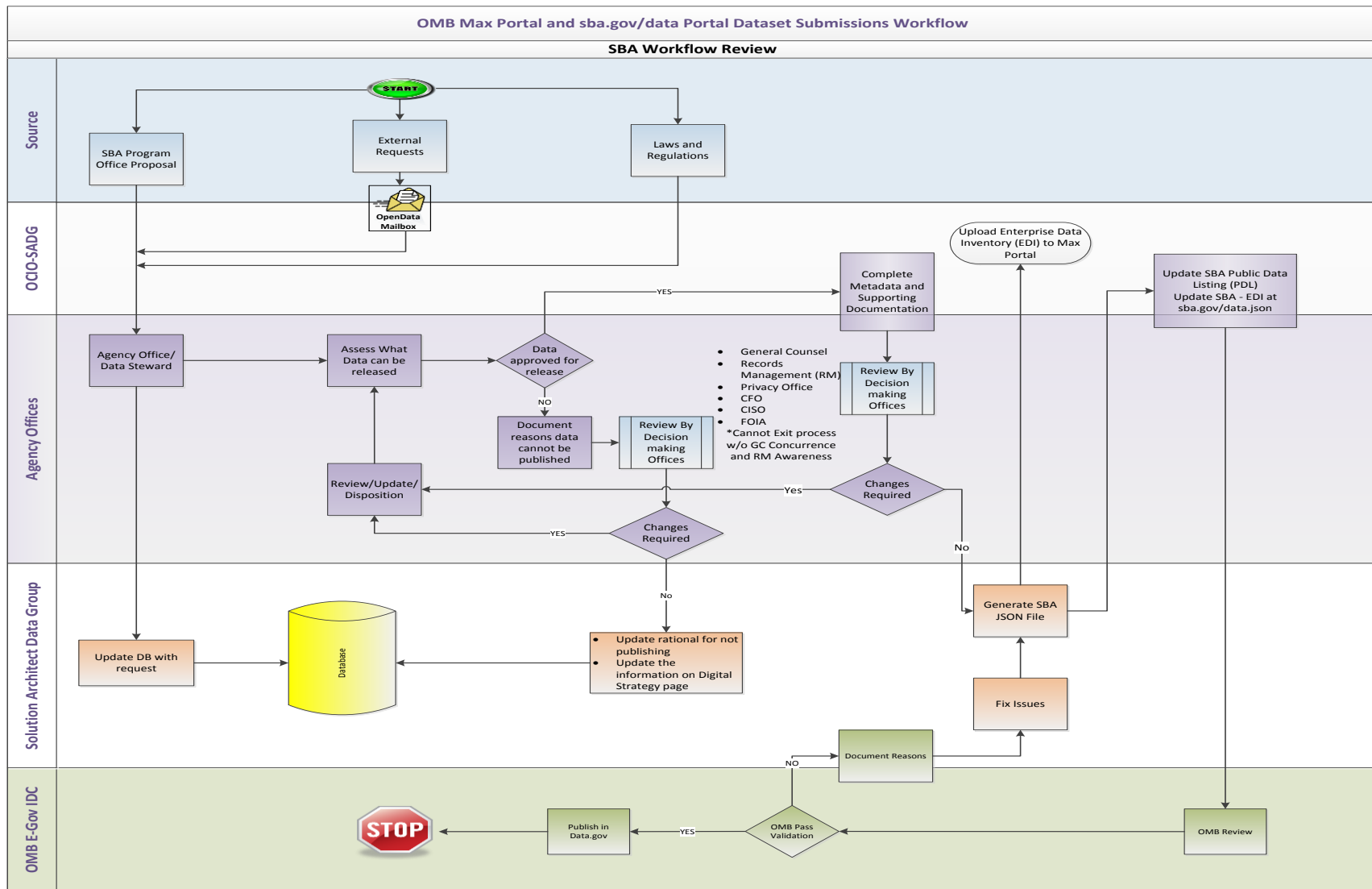
OMB Max Portal and sba.gov/data Portal Dataset Submissions Workflow

SBA Workflow Review

Source

START

SBA Program Office Proposal

External Requests

Laws and Regulations

OpenData Mailbox

OCIO-SADG

Complete Metadata and Supporting Documentation

Upload Enterprise Data Inventory (EDI) to Max Portal

Update SBA Public Data Listing (PDL) Update SBA - EDI at sba.gov/data.json

Agency Offices

Agency Office/ Data Steward

Assess What Data can be released

Data approved for release

YES

NO

Document reasons data cannot be published

Review By Decision making Offices

- General Counsel
- Records Management (RM)
- Privacy Office
- CFO
- CISO
- FOIA
*Cannot Exit process w/o GC Concurrence and RM Awareness

Review By Decision making Offices

Changes Required

Yes

No

Review/Update/ Disposition

Changes Required

YES

No

Solution Architect Data Group

Update DB with request

Database

- Update rational for not publishing
- Update the information on Digital Strategy page

Generate SBA JSON File

Fix Issues

Document Reasons

NO

OMB E-Gov IDC

STOP

Publish in Data.gov

YES

OMB Pass Validation

OMB Review

Figure 1: Workflow Overview

8

The workflow is broken down into five swim lanes: (Refer Figure 1: Workflow Overview)

- **Source** of requests for dataset publication
- **OCIO-SADG** team provides technical and administrative support for Open Data at SBA.
- **Agency Offices** are the custodians of data at SBA.
- **Solutions Architect Data Group** (SADG) provides technical support to Agency Offices.
- **OMB E-Gov. IDC** is responsible for collecting data for the Open Data Initiative.

The workflow, including responsibilities and tasks for each swim lane, is detailed below:

A. Request for Data (Source Swim Lane)

Multiple sources may generate requests for new or improved datasets to be published. Sources for these requests include:

- External Sources: Inquiries for data coming to [OpenDataFeedback@sba.gov.](mailto:OpenDataFeedback@sba.gov) They include requests from:
    - o The public
    - o State or Local Entities

These do not include FOIA/PA requests. Freedom of Information Act requests should be submitted to foia@sba.gov and Privacy Act requests should be mailed to FOI/PA Office, SBA, 409 Third St., SW, Washington DC 20416.

- Laws and Regulations: passed by Congress or OMB and related to Open Data Policy.
- Internal Sources: Agency Offices initiating the process to publish data.

Once the data request is received by the OCIO data group, it will be logged into the IEI Repository Database for tracking. An email request will be sent to the Data Steward for the relevant Agency Office (see Section Agency Office Roles (Agency Offices Swim Lane)).

B. OCIO Role (OCIO-SADG Swim Lane)

The OCIO team must provide support to Agency Offices throughout the process of reviewing and readying datasets for publishing. The OCIO-SADG must be responsible for uploading datasets to the Max Portal for publication.

For incoming data publication requests, the OCIO Technical Team's workflow includes:

- External requests received via the OpenDatafeedback@sba.gov mailbox. The OCIO Technical Team must analyze these requests and forward to the applicable Agency Offices.

Once it has been determined that a dataset can be released to the public, the OCIO team must assist Agency Offices with the publication process that includes:

- The completion of metadata and supporting documentation that will be saved in a repository (such as a Microsoft SharePoint site). The final information package will be updated in the IEI Repository Database. The list of required documentation is located in Section VI.F,

Required Artifacts.
- Publication of their datasets and information to the sba.gov/data portal and the corresponding SBA Public Data Listing (PDL) to http://www.sba.gov/data.json, and future standards.
- Upload of the Enterprise Data Inventory (EDI) to the OMB Max Portal:
  - The OCIO must ensure that all datasets are in required machine- readable format and satisfy OMB requirements before release.
  - The OCIO must publish the EDI – consisting of the JSON file, as well as any metadata and supporting documentation to the Max Portal.
  - Submissions to the E-Gov. IDC Max Portal require a username and password; the OCIO must be responsible for these credentials.
- Updates to the EDI, PDL, and the OMB Max Portal, with any additional or updated data (including metadata and supporting documentation).

C. Agency Office Roles (Agency Offices Swim Lane)

Agency Offices are the custodians and authoritative sources of data at SBA and are responsible for managing their data to maintain its relevance, quality, and privacy. Each Agency Office must assign a Data Steward who is the designated point of contact for Open Data. Agency Offices are responsible for obtaining approval of dataset publication and for following the process, with assistance from the OCIO - SADG, for publishing those datasets.

The Agency Office is responsible for initiating and/or responding to requests for data publication, as well as consulting with SBA stakeholders to approve data publication that includes:

- Responses to requests for dataset publication that are received by the Agency Office, either forwarded by the OCIO Technical Team or generated internally within the Agency Office, in order to comply with OMB directives, and/or meet legal and/or regulatory requirements.
- Assessments as to what data can be released and provide recommendations and justifications based on their analysis.
  - Agency Offices are responsible for consulting with the CFO, CISO and Freedom of Information/Privacy Acts Office for their recommendations in publishing or withholding data.
  - Agency Offices must get concurrence from OGC and notify Records Management.
  - If Agency Offices decide to release data, they must complete metadata and supporting documentation with the assistance of the OCIO Technical Team (refer to Section "VI.F Required Artifacts"), or else document the reasons for not submitting these additional materials, and send for approval to the appropriate decision-making offices.
  - If the data is approved for publication, the Agency Office coordinates with the OCIO-SADG to prepare the data for public distribution.
  - If the data is not approved for publication, a change or review is requested. Then the data approval process is repeated.
  - If the request is denied, then the Agency Office must document the reason and

submit the required documentation to OCIO for publication on the SBA Digital Strategy page. This information will also be recorded in the IEI Repository Database.

Agency Offices are responsible for the process of data publishing and must be supported by the OCIO. These tasks include:

- Work with the OCIO team to publish the machine-readable metadata file (Ex. JSON files) for the dataset that is published.
- Coordination with the OCIO-SADG for publishing data to the PDL, EDI, and OMB's Max Portal.

D. Solution Architect Data Group (SADG) Activities (Solution Architect Data Group Swim Lane)

The OCIO-SADG must provide technical support to the Agency Offices and publish datasets. The OCIO is the primary contact for the OMB with respect to management of Open Data. The responsibilities of the SADG must include:

- Management of SBA Open Data Database.
- Assistance to Agency Offices in compiling metadata and supporting documentation for all publicly-releasable data that is approved for publishing.
    - o A repository such as Microsoft SharePoint site will be used as a data collection location.
- Assistance to the Agency Office in packaging the data for publishing, that includes generating a standard SBA identifier for the dataset and creating the machine-readable format standard file.
- Publication of rationales for withheld datasets to the Digital Strategy page.
- SBA will publish a general overview of this process on the www.SBA.gov/digitalstrategy page. This must include process by which data is determined to be non-releasable, and examples of data characteristics that lead to 'No Release' determination.
- Coordination with OMB to comply with Open Data requirements. The group works closely with the OMB team in addressing issues related to publishing data to the SBA Internal Portal, the OMB Max Portal, and Data.Gov.

E. Data Validation and Publishing (OMB E-Gov. Integrated Data Collection Compliance Reporting Swim Lane)

OCIO-SADG is responsible for coordinating with OMB through the process of reviewing and publishing datasets, metadata and supporting documentation.

 OMB Compliance Reporting will include:

- Review of all submitted deliverables for SBA
- Validation and reporting of the findings
- Resolution of any identified issues
- Publication of SBA data to Data.gov

F. Required Artifacts

The following standard documents are needed to support the review and approval process at SBA. As the review proceeds, these documents may be requested by decision-making offices. The OCIO will assist Agency Offices in compiling these standard documents. All these required artifacts and documents, including the signed Form 58, should be maintained by the Agency Office/System Owner as part of the official System Documentation.

1. Metadata Template

This document is required for submission to the OMB Max Portal and provides the OMB team with the necessary information to properly register the dataset on Data.gov. This document must also be used in SBA's publication process for the EDI and PDL. At present, OMB accepts a "JSON" metadata file format. Please refer to Appendix E: JSON Metadata Template for details.

2. SBA Internal Enterprise Inventory(IEI) Metadata and Supporting Documentation

Depending on the data being published, the following documents are necessary for discoverability of the data:

- Data dictionary
- Logical and physical data models
- XML schema
- Directions for using the web service
- Source code
- Data samples

Like the data itself, supporting documentation must be approved for publication by decision-making offices. These documents will be stored in a centralized location on SharePoint that is accessible to all Agency Offices, the OCIO Technical Team and the Solution Architecture Data

Group. The SADG will build and maintain a central SBA Internal Enterprise Inventory (IEI) Repository Database to house all the artifacts for each dataset.

To promote data interoperability and openness, all Agency Offices in SBA must use the Internal Enterprise Inventory (IEI) as an open data standard repository for information collection and creation. All new systems, system updates, and new acquisitions must reference the IEI. Request for changes from the existing data model will require a review and consent from the CIO. SBA's data steward and/or Program Managers must use these metadata specifications and review them for compliance with the common core metadata standard, specifications, and formats stored in the IEI.

3. *Form 58*

SBA Form 58 is a required sign-off form that is used by the General Counsel and other stakeholders as needed and by Agency Offices for the approval of datasets that need to be published. Please refer to Appendix B: SBA Form 58

Dataset Documentation

Name	Purpose	Required	Location
Metadata Template	Data.gov submission	Yes	Please see: http://www.sba.gov/data.json
Form 58	Sign-off sheet for General Counsel and Agency Office members	Yes	A sample form is displayed in "SBA Form 58".
Internal Enterprise Inventory (IEI) Metadata and Supporting Documentation	Documents to help understand the data (data dictionary, physical and logical models, XML schema, directions for using the web service, source code, and data samples)	Yes	n/a

Appendix A: Definitions

Data: The term "data" refers to all structured information, unless otherwise noted.

Data Dictionary: A set of information describing the contents, format, and structure of a database and the relationship between its elements. It is used to control access to and manipulation of the database.

Data Steward: Data Steward is the Agency office representative who will be the point of contact for managing data submissions and ensuring that it is current. A Data Steward's responsibility stems from an understanding of the business domain and the interaction of business processes with data. Data stewards work with system owners, data administrators (if required), and other staff to plan and execute an office-wide data governance, control, and compliance.

Dataset: The term "dataset" refers to a collection of data presented in tabular or non-tabular form.

Government information: "Government information" means information created, collected, processed, disseminated, or disposed of, by or for the Federal Government.

Information life cycle: The term "information life cycle" means the stages through which information passes, typically characterized as creation or collection, processing, dissemination, use, storage, and disposition.

Information: The term "information" means any communication or representation of knowledge such as facts, data, or opinions in any medium or form, including textual, numerical, graphic, cartographic, narrative, or audiovisual forms.

Internal Enterprise Inventory (IEI): An Internal Enterprise Inventory is a repository that identifies each dataset published by SBA under Open Data. It contains for each dataset the metadata of every data element being released which is extracted from the existing logical and physical schemas of the data asset.

Logical Data Model: "Logical data models" represent the abstract structure of a domain of information. They are often diagrammatic in nature and are most typically used in business processes that seek to capture items of importance to an organization and how they relate to one another.

Metadata: A set of data that describes and gives information about the data.

Open Data: The term "open data" refers to publicly-available data structured in a way that enables the data to be fully discoverable and usable by end users. In general, open data will be consistent with the following principles:

- *Public*. Consistent with OMB's Open Government Directive, agencies must adopt a presumption in favor of openness to the extent permitted by law and subject to privacy, confidentiality, security, or other valid restrictions.
- *Accessible*. Open data are made available in convenient, modifiable, and open formats that can be retrieved, downloaded, indexed, and searched. Formats should be machine-readable (i.e., data are reasonably structured to allow automated processing). Open data structures do not discriminate against any person or group of persons and should be made available to the widest range of users for the widest range of purposes, often by providing data in multiple formats for consumption. To the extent permitted by law, these formats should be non-proprietary, publicly-available, and no restrictions should be placed upon their use.
- *Described*. Open data are described fully so that consumers of the data have sufficient information to understand their strengths, weaknesses, analytical limitations, security requirements, as well as how to process them. This involves the use of robust, granular metadata (i.e., fields or elements that describe data), thorough documentation of data elements, data dictionaries, and, if applicable, additional descriptions of the purpose of the collection, the population of interest, the characteristics of the sample, and the method of data collection.
- *Reusable*. Open data are made available under an open license that places no restrictions on their use.
- *Complete*. Open data are published in primary forms (i.e., as collected at the source), with the finest possible level of granularity that is practicable and permitted by law and other requirements. Derived or aggregate open data should also be published but must reference the primary data.
- *Timely*. Open data are made available as quickly as necessary to preserve the value of the data. Frequency of release should account for key audiences and downstream needs.
- *Managed* Post-Release. A point of contact must be designated to assist with data use and to respond to complaints about adherence to these open data requirements.

Personally identifiable information: "Personally identifiable information" (PII) refers to information that can be used to distinguish or trace back to an individual' s identity, either alone or when combined with other personal or identifying information that is linked or linkable to a specific individual. The definition of PII is not anchored to any single category of information or technology. Rather, it requires a case-by-case assessment of the specific risk by which an individual can be identified. In performing this assessment, it is important for an agency to recognize that non-PII can become PII whenever additional information is made publicly-available (in any medium and from any source) that, when combined with other available information, could be used to identify an individual.

Physical Data Model: A "Physical Data Model" is a graphical representation of a physical database design. It is designed with the capabilities of a specific database management system and optimizes the implementation of detailed data requirements and business rules that reflect technology constraints, application usage, and performance requirements.

Privacy: refers to protecting any Agency information that is not releasable to the public.

Quality: Quality refers to maintaining timeliness of data, state of completeness, validity, consistency, timeliness and accuracy, thus increasing its usability.

Solution Architecture Data Group (SADG): OCIO's team assigned to provide technical help to the various Agency Offices in addition to coordinating with the OCIO team and OMB, as necessary, to publish the datasets.

Structured Data: Data that resides in fixed fields within a record or file. Relational databases and spreadsheets are examples of structured data. Although data in XML and JSON files are not fixed in location like traditional database records, they are nevertheless structured, because the data are tagged and can be accurately identified.

Structured information is to be contrasted with unstructured information (commonly referred to as "content") such as press releases and fact sheets. As described in the Digital Government Strategy, content may be converted to a structured format and treated as data. For example, a web-based fact sheet may be broken into the following component data pieces: the title, body text, images, and related links. Data that resides in a fixed field within a record or file is called structured data. This includes data contained in relational databases and spreadsheets.

Unstructured Data: Data that is more free-form, such as multimedia files, images, sound files, or unstructured text. Unstructured data does not necessarily follow any format or hierarchical sequence, nor does it follow any relational rules. Unstructured data refers to masses of (usually) computerized information which do not have a data structure which is easily readable by a machine. Examples of unstructured data may include audio, video and unstructured text such as the body of an email or word processor document.

Appendix B: SBA Form 58

U.S. SMALL BUSINESS ADMINISTRATION RECORD OF CLEARANCE AND APPROVAL	IDENTIFICATION NO.

SUBJECT
DATA.GOV DATA SUBMISSION

PERSON TO BE CONTACTED ON ATTACHED

NAME	OFFICE	PHONE
JANE Q. PUBLIC	OCA/OFA	205-0000

EXPLANATION
Request to submit data x to data.gov. This data represents the number of annual approved loans by program area and county.

RELEASED FOR CLEARANCE AND DISTRIBUTION
If this document is to be distributed electronically, I certify that the accompanying disk contains only one WordPerfect file, which is identical to the official signed document attached.

TITLE	SIGNATURE	DATE	DEADLINE DATE FOR CLEARANCE

CONCURRENCE IS ASSUMED IF NOT RETURNED BY THE DEADLINE DATE

OFFICE/OFFICIAL	SIGNATURE	DATE	CONCUR No Comment	CONCUR Comment	NON-CONCUR	COMMENT ADOPTED
Chief Operating Officer			☐	☐	☐	☐
Chief Privacy Officer			☐	☐	☐	☐
FOIA Representative			☐	☐	☐	☐
General Counsel Rep.			☐	☐	☐	☐
Dir., Off. of Performance Mgmt			☐	☐	☐	☐
			☐	☐	☐	☐
			☐	☐	☐	☐
			☐	☐	☐	☐
			☐	☐	☐	☐
			☐	☐	☐	☐
			☐	☐	☐	☐
			☐	☐	☐	☐
			☐	☐	☐	☐
			☐	☐	☐	☐

APPROVED (TITLE)		ROOM NO.
		RETURN TO ▬▬▬

SBA Form 58 (7-95)Use 9-94 Edition Until Exhausted

Appendix C: Requirements Checklist

Requirements	Checklist
1. Identify datasets for release to the public. Inform SADG/OCIO to record information.	☐
2. Collect Metadata information about the system. (Physical, logical models, Data dictionary)	☐
3. All Agency Offices will need to determine which data can/cannot be released and how often they will refresh the data to keep it up-to-date.	☐
4. After determining what data will be released, developers need to be engaged to create scripts that will generate data. The data must be stored in a machine-readable format like "JSON", "XML", "csv", etc.	☐
5. Engage with the Solution Architecture Data Group to identify a location for publishing the datasets. This may involve designing web pages to store the data and metadata.	☐
6. Creating a data listing of information that can be released to the public. This information needs to be updated on: http://www.sba.gov/data	☐
7. Update the JSON file in the Public Data Listing (PDL) located at: http://www.sba.gov/data.json Agency Offices need to provide information about the title, description, keyword, contact point, contact mailbox, access URL, etc. to the SADG team. The full list of information is available in the JSON file. This information will be used to update the PDL.	☐
The Solution Architect Data Group (SADG) will provide a unique identifier for each dataset.	☐
8. Update the Enterprise Data Inventory (EDI) on OMB Max Portal at http://www.max.omb.gov Once Agency Offices approve the data release, the OCIO Technical Team will upload the same JSON file published in the PDL to the MAX Portal.	☐
9. If Agency Offices cannot release any data, they must provide documentation of the reasons for not publishing. Agency Offices will need to submit the information to the SADG. They must post the information in the Digital Strategy page at: http://www.sba.gov/digitalstrategy	☐

Appendix D: RACI Chart

Project:	Open DATA
PROCEDURE	Open Data Publication
DEPARTMENT	All Departments
UPDATED	4/18/2014

Steps will be highlighted if A and R are not assigned.

R: Responsible for doing the step.

A: Accountable for the step.

C: Consulted with before the step.

I: Informed when the step is completed.

STEP	Description	OCIO Tech. Team	Program Office	Solution Architect Data Group	OMB
1	Review external requests and forward, as appropriate, to the applicable Agency Office.	A/R		C/I	
2	Agency Offices review internal and external requests.	I	A/R	C/I	
3	Respond to any requests for dataset publication that are received by the Agency Office.	C	A/R	A	
4	Assess what data can be released and provide recommendations and justifications based on their analysis.		A/R		

STEP	Description	OCIO Tech. Team	Program Office	Solution Architect Data Group	OMB
5	Consult with the General Counsel, Records Management, Privacy Office, CFO, and CISO, FOIA, etc., for their recommendations on publishing or withholding data.		A/R	I	
6	If the request is ultimately denied, a reason is published on the SBA Digital Strategy page and recorded in the database.		C	A/R	
7	Compile the metadata and supporting documentation.	C	A/R	C/I	
8	Generate a standard SBA identifier for the dataset and create the JSON file.	I	C/I	A/R	
9	Upload Enterprise Data Inventory to the Max Portal.	A/R	I	C	
10	Address any issues related to publishing data to the SBA Internal Portal, OMB Max Portal, and Data.Gov.		C/I	A/R	A/R
11	Update the SBA Public Data Listing (PDL).	A/R	I	C	
12	Update the SBA EDI at sba.gov/data.json.	A/R	C/I	C	
13	Review all submitted deliverables for SBA.			C/I	A/R
14	Validate and report findings.			C/I	A/R
15	Resolve any identified issues.		C/I	A/R	C
16	Publish SBA data to Data.gov.			I	A/R

Appendix E: JSON Metadata Template

Tag	Description	Required	Comments
title	Dataset name	Yes	
description	Any information describing the dataset	Yes	
keyword	Search keywords	Yes	Use keywords for improving search
modified	Date Modified	Yes	
publisher	Publisher of data	Yes	"U.S. Small Business Administration"
contactPoint	Name of contact person	Yes	Agency Office/ Data owner name
mbox	Email address of contact person	Yes	
identifier	Unique Identifier assigned for the dataset	Yes	SADG assigns the unique identifier
accessLevel	Access level	Yes	Ex: "public"
bureauCode	Agency's Bureau code	Yes	For SBA: ["028:00"]
programCode	Program code	Yes	Example: ["028:001"]
accessLevelComment	Comments for AccessLevel	Yes	If you don't have a comment, pass null
accessURL	URL of the dataset	Yes	
format	The open data format of the dataset	Yes	"xml", "json", etc
theme	Type of data being published	Yes	
systemofRecords	System from which data originated	Yes	