

GSA ORDER

SUBJECT: GSA Information and Data Quality Handbook

1. Purpose. This order issues and transmits Handbook (HB), General Services Administration (GSA) Information and Data Quality Guidelines.
2. Applicability. This order applies to all GSA employees involved in information and data management.
3. Background. The intent of this handbook is to develop a framework for consistent information and data set management methods. Formally establishing information and data quality principles allows GSA to robustly leverage data for its own business use as well as sharing it with the public. Section 515 of the Treasury and General Government Appropriations Act of 2001 mandates that GSA maximizes the quality, objectivity, utility, and integrity of information it disseminates. And H.R.4174 - Foundations for Evidence-Based Policymaking Act of 2018, Title II, "OPEN Government Data Act," requires open government data assets made available by federal agencies to be published as machine-readable data.
4. Explanation of Changes.
  - a. Added reference to H.R.4174 - Foundations for Evidence-Based Policymaking Act of 2018, Title II, "OPEN Government Data Act," to the background section;
  - b. Added the name of the current CIO to the signature block;
  - c. Updated definitions and tables and removed outdated links and references;
  - d. Replaced the term "data supplier" with "data steward";
  - e. Added a section for the role of data architect; and
  - f. Updated internal strategic goals.

/S/ \_\_\_\_\_

DAVID SHIVE  
Chief Information Officer  
Office of GSA IT

## TABLE OF CONTENTS

| Chapter Number and Description                                      | Page Number |
|---|-------------|
| <a href="#">Chapter 1 – Information and Data Quality Guidelines</a> | 3           |
| <a href="#">Chapter 2 – GSA and the Data Quality Lifecycle</a>      | 9           |
| <b>Appendix</b>   |             |
| <a href="#">A: DATA.gov Submissions</a>                             | 12          |
| <a href="#">B: Glossary of Relevant Terms</a>                       | 20          |
| <a href="#">C: References</a>                                       | 22          |

### CHAPTER 1 – INFORMATION AND DATA QUALITY GUIDELINES

1. Purpose. This handbook establishes information and data quality guidelines that can be applied for consistent information reporting, sharing and exchange. The scope includes identifying principles that present a high-level description of key components needed for effective enterprise-wide information and data management.

2. Background. Identification of mission-critical information is essential to fulfilling organizational business processes and objectives. Quality information serves as an enabler for cooperation, interoperability, and strategic decision making. Establishing quality guidelines emphasizes the development of internal processes that facilitate consistency and helps users create value from information and data.

3. Data Quality. Data quality entails the accuracy, accessibility, timeliness, completeness, consistency, and granularity throughout the lifecycle from point of origination to point of consumption. Information quality should include establishing plans for accurately capturing, processing, modeling, sharing, distributing, securing, presenting and operationalizing data and data products in order to enable data driven decisions. Managing data quality must include selecting and defining relevant quality measures within the appropriate business context. The enterprise data management framework should provide a logical, model-driven approach to improve quality and overcome barriers to interoperability by:

a. Providing a comprehensive model framework that describes common business entities and relations. This framework should also incorporate business rules for data validation, coordination, and integration.

b. Employing a common vocabulary that is used to translate among diverse data formats in order to maintain consistency in the meaning of terms.

c. Integrating OMB Federal Enterprise Architecture (FEA) Data Reference Model (DRM) concepts to enhance organizational capability to discover and reuse federated

data.

d. Using standards based exchange formats to describe metadata.

e. Providing a business modernization blueprint as a methodology for modeling processes and identifying shared services.

**Table 1.3.1 Data Quality Descriptions**

| <b>Term</b>                   | <b>Description</b>  |
|-------------------------------|---|
| Accessibility                 | The ease access for authorized users and ensure controls to prevent unauthorized access. Examples include but are not limited to: <ul style="list-style-type: none"><li>- Robust data catalogue</li><li>- Section 508 of the Rehabilitation Act of 1973 Compliance</li></ul>        |
| Accuracy                      | The degree to which data correctly reflects the real-world object or event being described. This includes a level of precision.   |
| Completeness                  | Within the universe of the given analysis, the volume of the data within that universe is available for analysis.   |
| Consistency                   | The degree to which two data instances provide correlating information about the same underlying object. The values should be consistent across data sets, and the interdependent attributes of these data sets should appropriately reflect their expected consistency.            |
| Data                          | Data relates to a fact, event, or transaction. (1) Data may become information when it is interpreted, modeled, and analyzed; and (2) Data can be multimodal including symbols, such as words (text and/or verbal), numbers, diagrams, images (still &/or video), and others, etc.. |
| Data & Information Life Cycle | The stages through which information passes, typically characterized as creation, collection, sourcing, ingestion, access, adsharing, integrating, analyzing and modeling, visualizing, presenting, and decision-making.  |
| Data Ingestion                | The process of collecting data which will be processed and used later to fulfil certain purposes. Ways of capturing data can range from high end technologies to low end paper instruments used in the field.   |
| Data Integrity                | In the context of data and information, data integrity describes a degree of trustworthiness.   |
| Data Model                    | The main aim of a data model is to support the development of information systems by providing the definition and format of data.   |

|                          |   |
|--------------------------|---|
| Data Set                 | A data set is an organized collection of data. The most basic representation of a data set is data elements presented in tabular form. A data set may also present information in a variety of non-tabular formats, such as an extended mark-up language (XML) file, a geospatial data file, or an image file, etc  |
| Distribution             | A function or a listing which shows all the possible values or intervals of the data with their associated weights.   |
| Granularity              | The level of detail in the data; number of attributes describing the data.  |
| Information              | The term "information" refers to meaning assigned to data through the process of interpretation, modeling, and analysis. In other words, information is data that has been processed in such a way as to be meaningful to the person who receives it.   |
| Information Architecture | The framework for organizing the planning and implementation of information resources. The set of information, processes, and technologies that an enterprise has selected for the creation and operation of information systems.   |
| Information Model        | Shows the relationships or linkages between major areas of interest to the business. It creates a shared business vocabulary (e.g. semantic model), defining a community's agreement on important concepts and relationships between those concepts.  |
| Metadata                 | Semantic information associated with a given variable includes business definitions of the data and clear, accurate descriptions of data types, potential values, original source system, data formats, and other characteristics. Metadata defines and describes business data. Examples of metadata include data element descriptions, data type descriptions, attribute/property descriptions, range/domain descriptions, and process/method descriptions according to the International Organization for Standardization (ISO) 11179-3. |
| Modeling                 | The process of transforming data into a form that is amenable to performing analysis and decision-making of either a whole information system or parts of it to communicate connections between data points and structures.   |
| Operationalizing         | Placing the data into a working order.  |
| Presenting               | Organizing the data into tables, graphs, or charts, so that logical and statistical conclusions can be derived from the collected measurements.   |
| Processing               | A series of operations on data, especially by a computer, to retrieve, transform, or classify information.  |

|            |  |
|------------|--|
| Securing   | The practice of protecting digital information from unauthorized access, corruption, or theft throughout its entire lifecycle.           |
| Sharing    | Allowing access to data by authorized users throughout the data lifecycle and across applications subject to applicable data governance. |
| Timeliness | Timeliness measures how quickly a change in data at the source propagates throughout the data system and at the point of consumption.    |

The topics mentioned above can be referenced on the enterprise architecture website located at:  
<https://www.gsa.gov/reference/reports/information-quality-guidelinessec-515/data-quality-guidelines>.

4. Data Stewardship. The Data Steward creates data and maintains its consistency and correctness within a given domain of responsibility. Each data set must have a steward responsible for the completeness, accuracy, security, and validity of the data, both at an individual and aggregate level within their domain. Stewards should work to guide the appropriate stakeholders in implementing the processes and policies necessary to support quality data.

a. Data stewardship is responsible for ensuring that consistent and cohesive procedures are in place for achieving the following objectives:

- (1) Identifying, administering, and coordinating the alignment of strategic goals to advance the business value of data sets.
- (2) Identifying datasets within domain definitions and access controls.
- (3) Identifying interrelations within other data sets.
- (4) Providing direction to the development and maintenance of data sets and ensuring that new and existing data sets are defined in a reliable manner.
- (5) Defining, assigning, and communicating initiatives involving version control, archiving, and data set decommissioning techniques.
- (6) Continuous quality improvement of data sets to maximize the value of information and increase customer satisfaction.

(7) Periodic review and update of metadata attributes whereby enabling end-users to search data sets with greater precision.

5. Data Architect. The Data Architect designs, creates, deploys, and manages an organization's data architecture. Data architects define how the data will be stored, consumed, integrated and managed by different data entities and IT systems, as well as any applications using or processing that data in some way. The data architect is responsible for achieving the following objectives among others:

- a. Data modeling and making it business relevant as well as establishing and maintaining naming standards
- b. Managing and disseminating metadata to facilitate the common understanding of data and encouraging its reuse
- c. Integrating/developing data ecosystem that connects data across disparate developing solutions
- d. Designing and developing big data technologies for onboarding structured and unstructured data
- e. Developing architectures that look at data throughout its lifecycle

6. Information Quality. Information quality must be realized through utility, objectivity, integrity, transparency, timeliness, and reproducibility. The information quality principles identified below communicate objectives and the creation of plans for the use and sharing of information across the organization. For sharing the data with the public, additional guidelines are to be considered; please refer to appendix A. As identified in the table below, the Information Quality Assessment Process (IQAP), along with its activities and sample work products should be used in analyzing information.

#### **IQAP Activities and Sample Work Products**

|   |   |   |
|---|---|---|
| <b><i>Determine Scope</i></b>             | Target Information Assets<br>Define Quality Metrics<br>Cost Analysis      | Information Quality<br>Improvement<br>Vision Document |
| <b><i>Analyze Information Quality</i></b> | Assess Logical and/or Physical Model<br>Assess Information Content        | Information Quality<br>Analysis Report                |
| <b><i>Implement Solution</i></b>          | Define Criteria<br>Analyze the Impact<br>Develop the Execution Milestones | Information Quality<br>Implementation<br>Plan         |

a. Determine Scope. Phase 1 should prioritize target information assets and establish quality metrics.

(1) Target Information Assets. At a minimum, the following activities should be conducted:

(a) Survey or interview program area stakeholders to elicit information quality issues and requirements;

(b) Map information quality issues and requirements to the IT Strategic Plan goals to prioritize areas for improvement; and

(c) Develop an information quality improvement plan that includes defining success measures.

(2) Define Quality Metrics. The information quality processes should define metrics applicable to each governed data element. In defining quality metrics the following should be considered:

(a) The value of the information;

(b) The cost of errors; and

(c) The cost of resources to improve the metrics.

(3) Cost Analysis. A comparative analysis of the costs of quality vs. non-quality information should include the following categories:

(a) Process Failure Costs. Costs incurred when a process cannot be accomplished because of missing, inaccurate, incomplete, invalid, or otherwise poor information;

(b) Cost of Rework. Costs associated to the time used to reconcile poor or failed information and/or work around processes;

(c) Opportunity Costs. Indirect costs associated with lost or missed opportunities; and

(d) Infrastructure Costs. Costs associated with developing and reusing databases and applications.



b. Analyze Information Quality - The goal of Phase 2 should be to create management processes that focus on preventing quality issues.

(1) Assess Logical and/or Physical Model - This activity should involve reviewing the applicable model(s) against business requirements and objectives.

(2) Assess Information Content and Applicable Data Governance - Ensure that all parties have a consistent understanding of the information that is being collected and shared.

c. Implement Solution - The following actions should be used when applying the information quality solution in Phase 3:

(1) Define Criteria - The objectives for the solution need to be clearly stated and should correspond with business goals.

(2) Analyze Impact and Complexity of Meeting the Criteria - Evaluate the solution to verify that it accomplishes the desired quality improvement without creating issues.

(3) Prioritize and Develop the Execution Milestones - Identify the schedule, roles and responsibilities, improvement procedures, and lastly the validation methods for applying an information quality solution.

## CHAPTER 2 – GSA AND THE DATA QUALITY LIFECYCLE

1. Purpose. This section outlines the data quality lifecycle that will allow GSA to continually improve its data, and as a result equip GSA to better support its operational, tactical, and strategic activities.

### 2. Data Management Roles.

a. Data management involves the cooperation and communication among the following entities:

(1) Data Stewards. Data stewards should work with data owners and managers to continually improve agency data. Responsibilities include defining data governance policies and advising data owners and managers on the implementation of those policies. They should serve as overall coordinators for enterprise data delivery.

(2) Data Owners. Data owners should reside in every business function throughout GSA. Data owners should work with the data stewards to carry the primary responsibility for defining data requirements. Data owners should control access to data as well as oversee changes to data definitions.

(3) Data Managers. Data managers should work closely with the data stewards and data owners to implement data governance policies. There can be multiple data managers throughout the organization. Data managers should be a conduit for improvement ideas.

(4) Data Users. Data users can be widely characterized as the class of people or processes that utilize data. Data users play a critical role in communicating how data is employed and how it can be improved. Data by itself, without context, has no value, but can become information when a user interprets it.

b. Sponsorship must include:

(1) Chief Information Officer (CIO). The CIO has responsibility for ensuring that data quality policies and processes align with IT Strategic Plan and business lines. When applicable, the CIO may designate a data management governance workgroup to establish and maintain information quality standards.

(2) Heads of Services and Staff Offices (HSSO). HSSOs are responsible for sponsoring management and technical personnel involved in information and data transparency initiatives. The information and data derived from HSSOs should serve as an enabler for cooperation, interoperability, and strategic decision making.

3. Data Quality Lifecycle. In addition to applying the enterprise-wide data quality framework referenced in Chapter 1, GSA should actively engage the principles identified in the data quality lifecycle as an ongoing data management initiative. At the highest level, this process must address the six major components shown in Figure 2.3.



**Figure 2.3 - Data Quality Improvement Lifecycle**

a. Assess. The data quality lifecycle should start with an initial enterprise level or system-by-system assessment of data repositories. The objective should be to analyze the data entries and the data manipulation processes to find the root cause of errors and to highlight improvement opportunities. The output of the assessment should yield an updated data dictionary, clearly defined relationships among data elements, and a roadmap on how the organization will normalize its data sets.

b. Plan. The data owner should determine which improvements have the most far-reaching benefits. After opportunities for improvement have been defined and finalized they should be prioritized, approved, funded, staffed, and scheduled.

c. Execute. The data owner should openly communicate all details of a proposed data quality improvement initiative. This coordination should involve all business users who access the data, the database administrators who are maintaining it, and the developers whose programs have built-in interfaces to the data.

d. Evaluate. The data owner and/or steward should monitor the implemented improvements and determine its effectiveness. They should take into consideration the cost, accuracy, and performance results. If deemed necessary, changes that are problematic should have the flexibility to be reversed with minimal disruption to the organization.

e. Adapt. Data quality improvements that have been tested, verified, and accepted must be announced to the entire organization before turning them into new standards, guidelines, or procedures.

g. Educate. The final phase is to disseminate information about the data quality improvements that have been implemented. Depending on the scope of the change,

education can be accomplished through the organization's intranet, an internal newsletter, or a broadcasted email to all stakeholders.

4. Data Security. Throughout its lifecycle, data must be protected in a manner that is defined by the policies and procedures of GSA. The Information Security Office (IS) should be contacted directly if at any point there is concern or questions regarding any aspect of GSA data, to include but not limited to topics in access control, processing, or handling instructions.

5. Data Sharing. All data sharing efforts within GSA must meet agency-wide statutory, regulatory, and security mandates.

a. Internal. To maximize its use of information, GSA must enable the frictionless-flow of data within the agency. Data should be made available in a timely and responsive manner subject to applicable data sharing agreements (refer to #4 in Appendix C) and in support of the following strategic goals:

(1) IT Modernization: Provide a modern/streamlined experience for employees.

(2) Data Management: Unleash the power of data to inform and drive decision-making.

(3) Customer Experience: Deliver service comparable to leading private sector companies.

b. External. GSA's default position will be to share information openly with the public while meeting existing requirements that protect the release of inappropriate data. The example guidelines illustrated in Appendix A identify criteria that will allow GSA to fully participate in DATA.gov submissions. As other external data sharing programs are formalized, additional processes must be customized according to unique considerations.

## APPENDIX A

### DATA.gov Submissions

1. Background. DATA.gov is an initiative to allow the public to easily find, download, and use data sets that are generated and held by the Federal government. Improving access to agencies' data can help foster innovation and fuel the knowledge economy as well as increase transparency. DATA.gov will enable the public to participate in research and discovery by providing data from which they can build applications and conduct independent analyses.

2. Purpose. This section identifies GSA specific procedures and guidelines that anyone involved in releasing data sets to DATA.gov (e.g. program offices, data owners/stewards) should follow.

3. Metadata. The data steward must describe a data set according to attributes formalized by the DATA.gov Program Management Office (PMO). Metadata will provide information about the context of the data collection, data set completeness, and other factors that might influence the utility of the data for a specific purpose. The most critical elements of metadata include:

- data descriptions;
- keywords;
- data sources;
- URLs of technical documentation; and
- security considerations.

The data steward should think both broadly and specifically when selecting keywords; the robustness of the text-based search capability will determine the extent to which users can find the data in which they are interested. The most current metadata template may be obtained from the GSA Point of Contact (POC) or be directly entered through the DATA.gov Data Management System (DMS) (<https://www.data.gov/>).

4. Impact Assessment. The primary objective of an impact assessment is to identify the potential consequences of releasing data sets to DATA.gov. The data steward must ensure that no adverse effects will occur as a result of making its data public. Where applicable, the impact assessment may consist of two parts: technical analysis followed by risk management. Program offices should continue to follow their regular routine for making information or data public to include however many levels of approval they deem necessary. If a data steward is not confident regarding the components of an impact assessment they should contact the respective organization responsible for establishing the policy and/or regulation.

The GSA internal clearance process for GSA data assets, Order 2164.1 CIO Internal Clearance Process for GSA Data Assets, must be adhered to in accordance with its instructions.

The GSA guidelines below should be thoroughly examined to determine the suitability of sharing candidate data sets:

a. Security - (Security Categorization, Risk Assessment, Certification and Accreditation) (CIO IT Security 06-30) - 9/11/2020.

[https://www.gsa.gov/cdnstatic/Managing\\_Enterprise\\_Cybersecurity\\_Risk\\_%5BCIO\\_IT\\_Security\\_06-30\\_Rev\\_18%5D\\_09-11-2020docx.pdf](https://www.gsa.gov/cdnstatic/Managing_Enterprise_Cybersecurity_Risk_%5BCIO_IT_Security_06-30_Rev_18%5D_09-11-2020docx.pdf).

b. Privacy – CIO 2231.1 GSA Data Release Policy  
(<https://www.gsa.gov/directive/gsa-data-release-policy>).

c. Freedom of Information Act (FOIA) - Exemptions  
(<https://www.gsa.gov/reference/freedom-of-information-act-foia>).

d. Legal - Laws and Legal Matters  
<https://www.gsa.gov/policy-regulations/policy/information-integrity-and-access/it-security-procedural-guides>.

5. Data Policy Statements. Program offices and data stewards should adhere to the default data policy guidelines described below prior to releasing any data to the public.

a. Public Information - data sets accessed through DATA.gov are confined to public information and must not contain National Security information as defined by statute and/or Executive Order, or other information/data that is protected by other statute, practice, or legal precedent.

b. Security - Information accessed through DATA.gov must be in compliance with the required confidentiality, integrity, and availability controls mandated by Federal Information Processing Standard (FIPS) 199 as promulgated by the National Institute of Standards and Technology (NIST) and the associated NIST publications supporting the Certification and Accreditation (C&A) process.

c. Privacy - Information accessed through DATA.gov must be in compliance with current privacy requirements including OMB Circular A-130. In particular, GSA is responsible for ensuring that the data sets accessed through DATA.gov have any required Privacy Impact Assessments or System of Records Notices (SORN) easily available on GSA.gov. Under no circumstances, should data contain Personal Identifiable Information (PII).

d. Data Quality and Retention - Information accessed through DATA.gov is subject to the Information Quality Act (P.L. 106-554). For data accessed through DATA.gov, GSA must confirm that the data being provided meets GSA's Information and Data Quality Guidelines.

As the authoritative source of the information, GSA retains version control of GSA data sets accessed through DATA.gov in compliance with record retention requirements outlined by the National Archives and Records Administration (NARA).

e. Secondary Use - Data accessed through DATA.gov do not, and should not, include controls over its end use. However, as the data owner or authoritative source for the data, GSA must retain version control of data sets accessed. Once the data have been downloaded, GSA cannot vouch for quality and timeliness. Furthermore, GSA cannot vouch for any analyses conducted with GSA data retrieved from DATA.gov.

6. Office of Information and Regulatory Affairs (OIRA) Guidance.

a. GSA retains responsibility as the authoritative source of data, including corrections and updates.

b. GSA retains responsibility for protection of personally identifiable information and records retention.

c. GSA program office attests:

- The data set is in compliance with applicable privacy, confidentiality, and other relevant statutes; and
- The data set is in compliance with agency Information and Data Quality Guidelines.

7. Information and Data Dissemination Criteria. In an effort to make data dissemination consistent, as well as to improve compliance with existing statutory responsibilities, the OCIO and its designated governance working group have created a data set quality checklist. The Data.gov Checklist in section 8 should be referenced to ensure that data sets being released outside the agency conform to GSA's criteria.

GSA's DATA.GOV data set Checklist



DATA.gov Submission Checklist

|  |                          |   |
|--|--------------------------|---|
| <p><b>Data Set Name &amp; Description:</b></p><br><br><p>-----</p> <p>a) <u>Data Source</u> –</p> <p>-----</p> <p>b) <u>System / Application</u> –</p> <p>-----</p> <p>c) <u>Target Audience</u> –</p>   |                          |   |
| <p><b>High-value* [ Y / N ]:</b></p> <p><input type="checkbox"/> <i>Increase agency accountability and responsiveness</i></p> <p><input type="checkbox"/> <i>Improve public knowledge of the agency and its operations</i></p> <p><input type="checkbox"/> <i>Further the core mission of the agency</i></p> <p><input type="checkbox"/> <i>Create or expands economic opportunity</i></p> <p><input type="checkbox"/> <i>Respond to need and demand as identified through public consultation</i></p> <p><input type="checkbox"/> <i>Other</i></p> <p>If YES, please provide a brief explanation <u>geared to the citizen reader</u> of why it is high-value:</p><br><br> |                          |   |
|  |                          |   |
| 1  | <input type="checkbox"/> | The data set is public information and does not contain National Security or other information/data protected by statute, Agency practice, legal precedent, or otherwise restricted by GSA.   |
| 2  | <input type="checkbox"/> | The data set complies with required confidentiality, integrity, and availability controls for GSA, thereby adhering to NIST and OMB guidance.   |
| 3  | <input type="checkbox"/> | The data set is in compliance with GSA and OMB privacy requirements.  |
| 4  | <input type="checkbox"/> | <p>The data owner (<i>signature below</i>) certifies the data set meets GSA's Information Quality and Data Guidelines (CIO P 2142.1) to include the following components defined in Table 1-5-1 of the Handbook.</p> <ul style="list-style-type: none"> <li>▪ Accessibility</li> <li>▪ Completeness</li> <li>▪ Consistency</li> <li>▪ Timeliness</li> <li>▪ Accuracy</li> </ul> |



|   |                          |   |
|---|--------------------------|---|
|   |                          | <i>Where applicable please identify means, mechanisms, or persons that conducted/attest to the assessment:</i>  |
| 5   | <input type="checkbox"/> | The data owner and/or program office is the authoritative* source for the data and manages versioning and record retention requirements.  |
| 6   | <input type="checkbox"/> | The data set does not include controls over its end use. In the metadata for the data set, the Agency citation should note that the data was obtained from DATA.gov and that the Federal Government cannot vouch for its analyses after being retrieved from DATA.gov.  |
| 7   | <input type="checkbox"/> | The data set is a product of the Federal Government, or the government has unrestricted rights of use. The data set must be suitable for listing or downloading through any of the DATA.gov catalogs. The "Raw" Data Catalog provides an instant download of machine readable, platform-independent data sets, while the "Tools" Catalog provides hyperlinks to web pages that allow mining and/or downloading of raw data. |
| 8   | <input type="checkbox"/> | RAW DATA CATALOG: The format of the data set is one of the following: XML, CSV/TXT, KML/KMZ, Excel (XLS), ESRI Shapefile or in another machine readable format. (Data in HTML and PDF files should NOT be considered for the "Raw" Data Catalog.)   |
| 9   | <input type="checkbox"/> | TOOLS CATALOG: If single or multiple raw data sets are offered within a "Tool" environment, the tool that offers the raw data set(s) is one of the following: (1) Data Extraction Tool or webpage with downloadable data sets; (2) Feeds such as RSS, Atom or CAP; (3) a Widget (anything that requires a login or restricts use of data is prohibited.)  |
| 10  | <input type="checkbox"/> | The data owner and/or program office understand they are responsible for hosting data submissions and that they should provide an active URL which DATA.gov will only reference (i.e., no data is uploaded directly to DATA.gov.)<br><br>-----<br><i>Please provide proposed Internet hosting location:</i>   |
| 11  | <input type="checkbox"/> | The data owner and/or program office agree to maintain the data set and respond to all public comments.   |
| 12  | <input type="checkbox"/> | The data owner and/or program office will submit updates to the data set, metadata, and necessary URL(s) in a timely manner.  |
| 13  | <input type="checkbox"/> | Complete and thoroughly describe the impact assessment of the proposed data set submission ( <i>Security, Privacy, FOIA, Legal</i> ):   |
| <p><i>* Authoritative data source is a recognized or official data production source with a designated mission statement or source/product to publish reliable and accurate data for subsequent users. An authoritative data source may be the functional combination of multiple or separate data sources.</i></p> <p><i>* High-value information is information that can be used to increase agency accountability and responsiveness; improve public knowledge of the agency and its operations; further the core mission of the agency; create economic opportunity; or respond to need and demand as identified through public consultation.</i></p> |                          |   |
| <b>Name: Requestor</b>  |                          | <b>Office Symbol</b>  |
| <b>Signature / Date</b>   |                          | <b>Position</b>   |
|   |                          | <b>Phone</b>  |
|   |                          | <b>E-mail</b>   |
| <b>Name: Data Owner (Management Accountability)</b>   |                          | <b>Office Symbol</b>  |
|   |                          | <b>Position</b>   |

|                                      |                      |                 |
|--------------------------------------|----------------------|-----------------|
| <b>Signature / Date</b>              | <b>Phone</b>         | <b>E-mail</b>   |
| <b>Name: OCIO Agency POC</b>         | <b>Office Symbol</b> | <b>Position</b> |
| <b>Signature / Date</b>              | <b>Phone</b>         | <b>E-mail</b>   |
| <b>Name: OCIO Executive Designee</b> | <b>Office Symbol</b> | <b>Position</b> |
| <b>Signature / Date</b>              | <b>Phone</b>         | <b>E-mail</b>   |

## 8. Roles and Responsibilities.

- a. The program office is responsible for determining which data sets and tools are suitable to be posted on DATA.gov.
- b. The program office retains the right and responsibility for managing its data and providing adequate technical documentation to include version control and archiving.
- c. The program office is responsible for ensuring that data stewards for a particular data set complete the required metadata.
- d. The data steward is responsible for ensuring that the data set is compliant with information and data quality guidelines in addition to completing an impact assessment.
- e. The program office, in conjunction with the data steward, is responsible for ensuring that their data sets are consistent with statutory responsibilities including those related to security, accessibility, privacy, and confidentiality.
- f. The program office and the CIO have the responsibility of ensuring that authoritative data sources are made available in formats that are platform independent and machine readable.
- g. The CIO has the responsibility for assigning an overall DATA.gov POC for the agency.
- h. The GSA POC is responsible for releasing data sets to DATA.gov along with facilitating comments back to program offices and data stewards.

9. DATA.gov Submission Process. The GSA DATA.gov submission process for candidate data sets to be published is described with regard to process scope, roles, decision criteria, and information flow. There are seven core process steps and three sub-processes that result from alternative decision paths. The process scope begins with any request for publication of a data set and ends with the publication of the data set on DATA.gov. The alternative decision path to end the process is to determine that GSA is not the authoritative source or the data set is not suitable for publication. The following paragraphs illustrate each process step.

- a. The GSA DATA.gov submission process begins when a customer requests publication of a data set. A customer can be internal or external to GSA. If the request is from a Government agency, the DATA.gov PMO will forward the request to the GSA POC. The GSA POC determines if GSA is the authoritative source for the requested data set. If it is, then the GSA POC identifies the appropriate data steward.
- b. The data steward determines if the data set is suitable for publication. Suitability is based on conformance to the information and data quality guidelines as well as GSA's DATA.gov checklist. If the data set is suitable, the data steward obtains approval of the

program office and completes the most current metadata template manually or through DMS.

c. The data steward conducts an impact assessment on the data set request package to test compliance with privacy, FOIA, legal, and security considerations. Upon a successful impact assessment, the data set proposal is forwarded to the GSA POC.

d. The GSA POC then reviews the completed document package and forwards it to the DATA.gov PMO for publication to the DATA.gov web site. Upon publication the core process is completed.

e. The alternative process is engaged under three separate conditions that occur at subsequent points in the process:

- First, if the initial request is a non-agency nomination and does not pass the initial screening and filtering requirements, the request is rejected and the data set is not published.
- Second, if the data set request is valid and passes the initial screening but it is determined by the GSA POC that GSA is not the authoritative source for the data set, the request would then be forwarded to the DATA.gov PMO to identify the correct authoritative source for the data set.
- Third, if the data set request is valid, and the agency is the authoritative source but the data steward deems the data set not suitable based on the information and data quality guidelines and/or impact assessment, the request is rejected and the data set is not published.

## APPENDIX B:

### Glossary of Relevant Terms

| Term                 | Description   |
|----------------------|---|
| Authoritative Source | An authoritative data source is a recognized or official data production source with a designated mission statement or source/product to publish reliable and accurate data for subsequent users. An authoritative data source may be the functional combination of multiple or separate data sources.  |
| Architecture         | Representation of the structure of a system or community that describes the constituents of the system and how they interact with each other such that the goals and responsibility of the system or community are met.   |
| Best Practice        | A group of tasks that optimizes the efficiency or effectiveness of the business discipline or process to which it contributes. Best practices are generally adaptable and replicable across similar organizations or enterprises - and sometimes across different functions or industries.  |
| Business Objective   | Objectives state what is to be achieved, and the results and activities required to measure progress towards reaching the desired state.  |
| Business Process     | <p>A business process is one aspect of a business model intended to specify the services, participants, interactions, resources and course of activities required to realize business value.</p> <p>A business process is a set of linked activities that create value by transforming an input into a more valuable output. Both input and output can be artifacts and/or information and the transformation can be performed by human actors, organizations, machines, or both.</p> <p>A business process can be decomposed into activities that may be atomic (E.G. "Delete file") or utilize sub-processes (E.G. "Build Ship"), which contributes to achieving the goal of the super-process. The analysis of business processes typically includes the mapping of processes and sub-processes down to activity level.</p> <p>A business process may specify how processes are currently executed or may specify a future-state process intended to improve business value and/or reduce costs.</p> |

|   |   |
|---|---|
| <p>Enterprise Architecture (EA)</p>                                     | <p>A process and the associated strategic information asset base to support enterprise objectives, which includes:</p> <ul style="list-style-type: none"> <li>- the mission;</li> <li>- information necessary to perform the mission;</li> <li>- the technologies necessary to perform the mission;</li> <li>- the transitional processes; and</li> <li>- the business model</li> </ul> <p>for supporting the transformation of an enterprise to meet its changing business objectives.</p>   |
| <p>Federal Enterprise Architecture (FEA) Data Reference Model (DRM)</p> | <p>The Data Reference Model (DRM) describes, at an aggregate level, the data and information supporting government program and business line operations. This model enables agencies to describe the types of interaction and exchanges occurring between the Federal government and citizens.</p> <p>The DRM categorizes government information into greater levels of detail. It also establishes a classification for Federal data and identifies duplicative data resources. A common data model will streamline information exchange processes within the Federal government and between government and external stakeholders.</p> <p>The DRM provides a standard means by which data may be described, categorized, and shared. These are reflected within each of the DRM's three standardization areas:</p> <ul style="list-style-type: none"> <li>- Data Description: Provides a means to uniformly describe data, thereby supporting its discovery and sharing</li> <li>- Data Context: Facilitates discovery of data through an approach to the categorization of data according to taxonomies; additionally, enables the definition of authoritative data sets within a community of interest (COI)</li> <li>- Data Sharing: Supports the access and exchange of data where access consists of ad-hoc requests (such as a query of a data set), and exchange consists of fixed, recurring transactions between parties</li> </ul> |
| <p>Information Technology (IT)</p>                                      | <p>The term 'information technology', with respect to an executive agency means any equipment or interconnected system or subsystem of equipment, that is used in the automatic acquisition, storage, manipulation,</p>   |

|                         |   |
|-------------------------|---|
|                         | management, movement, control, display, switching, interchange, transmission, or reception of data or information by the executive agency.  |
| Metric                  | A standard for measurement.   |
| Objectivity             | Involves a focus on ensuring that information is accurate, reliable and unbiased and that information products are presented in an accurate, clear, complete and unbiased manner  |
| Reproducibility         | Means that the information is capable of being substantially reproduced, subject to an acceptable degree of imprecision.  |
| Service or Staff Office | A Program Office within GSA responsible for coordinating nationwide programs and supporting Federal agencies and citizen-oriented organizations.  |
| Transparency            | A quality or characteristic of data or information. Transparency Promotes accountability and provides information for citizens about what their Government is doing. Transparency can strengthen the connections between government agencies and the public they serve. Transparency helps ensure meaningful and informed public participation. |
| Utility                 | The usefulness of the information to the intended users. The data provider should stay informed of changing information needs and develop new data, models, and information products where appropriate.   |

## APPENDIX C:

### References

1. Federal Enterprise Architecture (FEA) -  
<https://obamawhitehouse.archives.gov/omb/e-gov/FEA>
2. GSA Enterprise Architecture Transition Strategy and Sequencing Plan  
<http://ea.gsa.gov/>
3. GSA Information Quality Guidelines-Section 515.  
<https://www.gsa.gov/reference/reports/information-quality-guidelines/data-quality-guidelines>
4. D2D Data Sharing Agreement Documentation  
<https://insite.gsa.gov/services-and-offices/staff-offices/office-of-gsa-it/gsa-it-organizations/office-of-the-deputy-cio/data-to-decisions-d2d-program/d2d-data-sharing-agreement-documentation?term=data>
5. OMB Circular A-130 Revised  
<https://www.cio.gov/policies-and-priorities/circular-a-130/>
6. 2164.1 CIO Internal Clearance Process for GSA Data Assets  
<https://insite.gsa.gov/directives-library/internal-clearance-process-for-gsa-data-assets-21641-cio?term=data.gov%20submission%20process>
7. [GSA CIO 2105.1D](#), Section 508: Managing Information and Communications Technology (ICT) for Individuals with Disabilities, January 7, 2019.
8. [GSA CIO 2110.4](#), Enterprise Architecture Policy. May 24, 2017.