

## Data Custodian Skill Assessment Survey

### Looking to build deeper expertise as a Data Custodian?

As part of the Data Stewardship Service, we are offering this **optional** skill assessment to help you get an understanding of the skills that will help succeed in your role as a Data Custodian as well as gauge your individual skill level. After completing this assessment you'll receive your results. For the areas in which you'd like to advance your skills, the Data Stewardship Service will provide you with learning path recommendations.

#### Skill Assessment Overview:

As you move through this assessment, you will first see a set of core data terms. Our team wants to gauge your familiarity with these terms since they are essential to your role. After completing the data terms section, you'll go through **five t skill topic areas** and rate your ability to complete the Data Custodian task areas on a scale of **1- 5**, with '1' being the least knowledgeable, to '5' being the most knowledgeable.

#### Skill Assessment Topic Areas:

1. Enhance and Create Data Flows
2. Metadata and Reference Data Management
3. Data Integration and Interoperability
4. Golden Record Data Management
5. Data Governance

Your answers to this skill assessment survey will be confidential only to you and the Service core team. Our core team will use your results to make customized training recommendations for you so that you can grow your skills as a Data Custodian. If you have any questions regarding the survey, please reach out to the CoP Core team at [datastewardshipp servicedevelopment@nasa.onmicrosoft.com](mailto:datastewardshipp servicedevelopment@nasa.onmicrosoft.com).

#### Paperwork Reduction Act Statement:

*This information collection meets the requirements of 44 U.S.C 3507, as amended by section 2 of the Paperwork Reduction Act of 1995. You do not need to answer these questions unless we display a valid Office of Management and Budget control number. The OMB control number for this information collection is 2700-0153 and it expires on 07/31/2024. We estimate that it will take about 10 minutes to read the instructions, gather the facts, and answer the questions. You may send comments on our time estimate above to [briana.hila@nasa.gov](mailto:briana.hila@nasa.gov). Send only comments relating to our time estimate to this address.*

#### Section 1:

## Data Terms

These questions are designed to ensure that as an agency we use the same definition of data terms so that we can work more seamlessly as a data community. If the data term and definition matches with your definition/understanding of the term, select "Yes." If it doesn't, select "No."

1. **Data Quality:** Data that is accurate, complete, consistent, reliable, up-to-date, and fit for a purpose. Is this the definition you use to define data quality in your day-to-day work.
2. **Data Domain:** A logical grouping of items, or data, of interest to the organization.
3. **Data Catalog:** A software application that creates an inventory of an organization's data assets
4. **Data Interoperability:** The ways in which data is formatted that allow diverse datasets to be merged or aggregated in meaningful ways.
5. **Metadata Management:** Deliberate, structured data about data.
6. **Reference Data:** The data used to classify and reference other data.
7. **Golden Record Management (also known as Master Data Management):** A software/tool that logs/manages golden records.
8. **Golden Records (also known as Master Data Records):** A single, well-defined version of all the data entities in an organizational ecosystem.
9. **Golden Records (also known as Master Data Records):** A single, well-defined version of all the data entities in an organizational ecosystem.
10. **Extract, Transform and Load (ETL):** A data integration process that combines data from multiple data sources into a single, consistent data store that is loaded into a data warehouse or other target system.
11. **Business Glossary:** A collection of data related terms described in clear language that everyone in an organization can understand.
12. **Data Dictionary:** A collection of names, definitions, and attributes about data elements that are being used or captured in a database or information system
13. **Data Integration:** A practice that comprises the practices, architectural techniques and tools for achieving the consistent access and delivery of data across the spectrum of data subject areas and data structure types in the enterprise to meet the data consumption requirements of all applications and business processes.
14. **Data Lineage/Data Flows:** The process of tracking the flow of data over time, providing a clear understanding of where the data originated, how it has changed, and its ultimate destination within the data pipeline.
15. **Data Storage and Archival:** The design, implementation, and support of stored and archived data to maximize its value by managing the availability of data throughout the data lifecycle and ensuring the integrity of data assets
16. **Data Profiling:** A technology for discovering and investigating data quality issues, such as duplication, lack of consistency, and lack of accuracy and completeness.
17. **Centralized System:** Retrieves, collects, and stores data in one place.

**18. Information Assets:** Information relevant to an enterprise’s business function, including captured and tacit knowledge of employees, customers or business partners; data and information stored in highly-structured databases; data and information stored in textual form and in less-structured databases such as messages, e-mail, workflow content and spreadsheets; information stored in digital and paper documents; purchased content; and public content from the Internet or other sources.

**Selection Options for questions 1-18:**

- Yes
- No

**Section 2:**

**Enhance and Create Data Flows**

As a Data Custodian, you need to be able to write and support the execution of technical requirements related to **enhancing and creating data flows** from source to target in application systems landscape to improve data quality.

To help you complete this section, you may need to reference these data term definitions :

- **Data Domain:** A logical grouping of items, or data, of interest to the organization
- **Data Quality:** Data that is accurate, complete, consistent, reliable, up-to-date, and fit for a purpose
- **Data Profiling:** A technology for discovering and investigating data quality issues, such as duplication, lack of consistency, and lack of accuracy and completeness.

**19.** On a scale of 1-5, how would you rate your ability to complete the following tasks pertaining to enhancing and creating data flows?

	<b>1:</b> I’m not aware of this task	<b>2:</b> I’m aware but need to learn	<b>3:</b> I’ve learned but need coaching	<b>4:</b> I can complete this task on my own	<b>5:</b> I can teach this task
I can conduct system enhancements to embed data quality rules within systems to ensure the data is complete, consistent, accurate, timely, valid and unique for data at rest					
I can conduct data profiling to improve levels of data quality for specific use cases (SQL, Python, R, etc)					

I can conduct statistical analysis and implement statistical techniques such as hypothesis testing, regression analysis, and outlier detection to identify data quality issues					
I can write technical requirements to add data quality rules to data as it moves from one system to another					

**Section 3:**

**Metadata and Reference Data Management**

As a Data Custodian, you need to be able to coordinate with data stewards and IT to support the technical implementation of **metadata and reference data management**.

To help you complete this section, you may need to reference these data term definitions:

- **Data Domain:** A logical grouping of items, or data, of interest to the organization
- **Metadata Management:** Deliberate, structured data about data
- **Reference Data Management:** The data used to classify and reference other data
- **Data Catalog:** A software application that creates an inventory of an organization's data assets
- **Business Glossary:** A collection of data related terms described in clear language that everyone in an organization can understand
- **Data Dictionary:** A collection of names, definitions, and attributes about data elements that are being used or captured in a database or information system

**20.** On a scale of 1-5, how would you rate your ability to complete the following tasks pertaining to metadata and reference data management?

	<b>1:</b> I'm not aware of this task	<b>2:</b> I'm aware but need to learn	<b>3:</b> I've learned but need coaching	<b>4:</b> I can complete this task on my own	<b>5:</b> I can teach this task
I can partner with data stewards and data owners to determine requirements for system maintenance for metadata and reference data					

management tooling (Ex. Data Catalogs, Business Glossaries, Data Dictionaries)					
I can conduct system enhancements and support enhancements related to metadata and reference data changes					

**Section 4:**

**Data Integration and Interoperability**

As a Data Custodian, you need to be able to coordinate with data stewards and IT to support the technical implementation of **data integration and interoperability**.

To help you complete this section, you may need to reference these data term definitions:

- **Data Domain:** A logical grouping of items, or data, of interest to the organization
- **Data Quality:** Data that is accurate, complete, consistent, reliable, up-to-date, and fit for a purpose
- **Data Integration:** A practice that comprises the practices, architectural techniques and tools for achieving the consistent access and delivery of data across the spectrum of data subject areas and data structure types in the enterprise to meet the data consumption requirements of all applications and business processes
- **Data Interoperability:** The ways in which data is formatted that allow diverse datasets to be merged or aggregated in meaningful ways

**21.** On a scale of 1-5, how would you rate your ability to complete the following tasks pertaining to data integration and data interoperability?

	<b>1:</b> I'm not aware of this task	<b>2:</b> I'm aware but need to learn	<b>3:</b> I've learned but need coaching	<b>4:</b> I can complete this task on my own	<b>5:</b> I can teach this task
I understand how data modeling can be used to improve consistency across systems so data can be integrated more easily					
I understand which common protocols are widely accepted and are well documented (Ex.					

HTTP and HTTPS) for exchanging data between different systems					
I can understand the standardization of practices, and am knowledgeable about open and standardized data formats that are easily readable by many different systems and applications (Ex. CSV, JSON, XML)					
I can write technical requirements to add data quality rules to data as it moves from one system to another					
I can use data profiling techniques to systematically analyze data to determine data integration requirements					

**Section 5:**

**Golden Record Data Management**

As a Data Custodian, you need to be able to coordinate with data stewards and IT to support the technical implementation of **golden record data management**.

To help you complete this section, you may need to reference these data term definitions:

- **Data Domain:** A logical grouping of items, or data, of interest to the organization
- **Golden Records (also known as Master Data Records):** A single, well-defined version of all the data entities in an organizational ecosystem
- **Golden Record Data Management (also known as Master Data Management):** A software/tool that logs/manages golden records
- **Extract, Transform and Load (ETL):** A data integration process that combines data from multiple data sources into a single, consistent data store that is loaded into a data warehouse or other target system
- **Data Lineage/Data Flows:** The process of tracking the flow of data over time, providing a clear understanding of where the data originated, how it has changed, and its ultimate destination within the data pipeline

22. On a scale of 1-5, how would you rate your ability to complete the following tasks pertaining to golden record data management?

	1: I'm not aware of this task	2: I'm aware but need to learn	3: I've learned but need coaching	4: I can complete this task on my own	5: I can teach this task
I can understand authoritative sourcing for Extracting, Transforming and Loading (ETL) and have awareness of the sources of truth for the data used to create a golden record					
I can conduct an impact analysis and understand data lineage (data flows) for my respective NASA Office, and the impact of technical changes					

**Section 6:**

**Data Governance**

As a Data Custodian, need to be able to support data stewards with implementing the technical changes and controls required to support implementation and embedding of **NASA and Federal data policies**.

To help you complete this section, you may need to reference these data term definitions:

- **Information Assets:** Information relevant to an enterprise's business function, including captured and tacit knowledge of employees, customers or business partners; data and information stored in highly-structured databases; data and information stored in textual form and in less-structured databases such as messages, e-mail, workflow content and spreadsheets; information stored in digital and paper documents; purchased content; and public content from the Internet or other sources

**23.** On a scale of 1-5, how would you rate your ability to complete the following tasks pertaining to data governance?

	1: I'm not aware of this task	2: I'm aware but need to learn	3: I've learned but need coaching	4: I can complete this task on my own	5: I can teach this task
I can understand and have awareness of how NASA and Federal data policies effect the security and encryption of					

information assets (Ex. NIST SP 800-53, FISMA, etc)					
I can maintain, manage, understand, and have the ability to document any exceptions to data policies					
I can effectively use my knowledge of NASA access controls (Ex. NASA Identify and Access Management system and role-based access controls, MFA, etc)					
I can shape technical requirements for the secure processing, storage, and retrieval of data					

**Section 7:**

**Additional Feedback**

Your feedback matters - please let us know if there are additional skills for Data Stewards we should include, any concerns or questions you would like to share with the team, or any other general feedback.

**24.** Please provide any additional feedback below:

Answer: \_\_\_\_\_