

Data Steward Skill Assessment Survey

Looking to build deeper expertise as a Data Steward?

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 Steward 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 **nine skill topic areas** and rate your ability to complete the Data Steward 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. Data Terms
2. Data Quality
3. Metadata Management
4. Reference Data Management
5. Golden Record Data Management
6. Document and Content Management
7. Data Integration and Interoperability
8. Data Storage and Operations
9. Data Stewardship
10. Data Security

**Note: All skills in this survey have been based on the Federal Data Skills Catalog*

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 Steward. If you have any questions regarding the survey, please reach out to the CoP Core team at **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. 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 Management:** Consists of the practices, architectural techniques, and tools for achieving consistent access to 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. Is this the definition you use to define data management in your day-to-day work?
- 2. Data Governance:** The specification of decision rights and an accountability framework to ensure the appropriate behavior in the valuation, creation, consumption and control of data and analytics. Is this the definition you use to define data governance in your day-to-day work?
- 3. Data Architecture:** Describes how data is managed from collection through to transformation, distribution, and consumption. It sets the blueprint for data and the way it flows through data storage systems. It is foundational to data processing operations and artificial intelligence (AI) applications. Is this the definition you use to define data architecture in your day-to-day work?
- 4. Data Integration:** 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. Is this the definition you use to define data integration in your day-to-day work?
- 5. 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?
- 6. Data Domain:** A logical grouping of items, or data, of interest to the organization. Is this the definition you use to define data domain in your day-to-day work?
- 7. Data Catalog:** A software application that creates an inventory of an organization's data assets. Is this the definition you use to define data catalog in your day-to-day work?
- 8. Data Glossary (also known as Business Glossary):** A collection of data related terms described in clear language that everyone in an organization can understand. Is this the definition you use to define data glossary in your day-to-day work?

9. **Data Dictionary:** A collection of names, definitions, and attributes about data elements that are being used or captured in a database or information system. Is this the definition you use to define data dictionary in your day-to-day work?
10. **Data Interoperability:** The ways in which data is formatted that allow diverse datasets to be merged or aggregated in meaningful ways. Is this the definition you use to define data interoperability in your day-to-day work?
11. **Data Asset:** Any entity that is comprised of data. It may be a system or application output file, database, document, or web page. It also includes a service that may be provided to access data from an application. Is this the definition you use to define data asset in your day-to-day work?
12. **Data Science:** Combines math and statistics, specialized programming, advanced analytics, artificial intelligence (AI), and machine learning with specific subject matter expertise to uncover actionable insights hidden in an organization's data. These insights can be used to guide decision making and strategic planning. Is this the definition you use to define data science in your day-to-day work?
13. **Data Owner:** Individual in charge of the data in a certain data domain. A data owner must guarantee that the information inside that domain is correctly maintained across various platforms and business processes. Is this the definition you use to define data owner in your day-to-day work?
14. **Data Product:** An application or tool that uses data to help businesses improve their decisions and processes. Is this the definition you use to define data product in your day-to-day work?
15. **Data Storage and Operations (DataOps):** A collaborative data management practice focused on improving the communication, integration and automation of data flows between data managers and data consumers across an organization. Goal is to deliver value faster by creating predictable delivery and change management of data, data models and related artifacts. DataOps uses technology to automate the design, deployment and management of data delivery with appropriate levels of governance, and it uses metadata to improve the usability and value of data in a dynamic environment. Is this the definition you use to define DataOps in your day-to-day work?
16. **Data Archival:** The process of collecting older data and moving it to a protected location so that it can be retrieved if needed in a data forensics investigation. Archives are distinct from backups. With data archiving, the information is moved to free up storage resources. With backups, working data is copied so that it can be restored in the event of a system failure or disaster. Is this the definition you use to define data archival in your day-to-day work?
17. **Data Pipeline:** A means of moving data from one place (the source) to a destination (such as a data warehouse). Along the way, data is transformed and optimized, arriving in a state that can be analyzed and used to develop business insights. Is this the definition you use to define data pipeline in your day-to-day work?
18. **Data Lineage:** 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. Is this the definition you use to define data lineage in your day-to-day work?

- 19. Data Profiling:** A technology for discovering and investigating data quality issues, such as duplication, lack of consistency, and lack of accuracy and completeness. Is this the definition you use to define data profiling in your day-to-day work?
- 20. Metadata Management:** Deliberate, structured data about data. Is this the definition you use to define metadata management in your day-to-day work?
- 21. Reference Data:** The data used to classify and reference other data. Is this the definition you use to define reference data in your day-to-day work?
- 22. Golden Record Management (also known as Master Data Management):** A software/tool that logs/manages golden records. Is this the definition you use to define golden record management in your day-to-day work?
- 23. Golden Records (also known as Master Data Records):** A single, well-defined version of all the data entities in an organizational ecosystem. Is this the definition you use to define golden records in your day-to-day work?
- 24. 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. Is this the definition you use to define ETL in your day-to-day work?
- 25. 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. Is this the definition you use to define information assets in your day-to-day work?
- 26. Authoritative Data Source:** A repository or system that contains identity information about an individual and is considered to be the primary or most reliable source for this information. In the case that two or more systems have mismatched or conflicting data, the identity information within the authoritative data source is considered to be the most accurate. Is this the definition you use to define authoritative data source in your day-to-day work?
- 27. Business Intelligence:** Those systems, tools and technologies used by organizations to manage, visualize and extract valuable insights from their raw business data. Is this the definition you use to define business intelligence in your day-to-day work?

Selection Options for questions 1-18:

- Yes
- No

Section 2:

Data Quality

As a Data Steward, you need to be able to establish data management standards relating to **data quality** in your data domain.

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

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

	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 use my data domain expertise to assess data accuracy and validity					
I can write data quality rules to ensure data is complete, consistent, accurate, timely and valid					
I can understand the data quality levels needed so the data can be used for a specific purpose					
I can clean, standardize, and transform data so that it is consistent and suitable for analysis and reporting					

Section 3:

Data Stewardship

As a Data Steward, you need to be able to support compliance efforts by assisting with the implementation of **NASA and Federal data policies**.

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 Steward Identification Matrix (DSIM):** The DSIM is a tool to systematically capture the data steward point of contact per data domain and per NASA office

29. On a scale of 1-5, how would you rate your ability to complete the following tasks pertaining for the following Data Stewardship competency area:

Data Stewardship Competency Area #1: Identifying Data Stewardship Roles

- *The following tasks assess your ability on identifying data stewardship roles across the service.*

	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 maintain entries in the Data Stewardship Identification Matrix (DSIM)					
I can define the data domains within my respective organization, and define data stewardship and the typical responsibilities of a Data Steward					

30. On a scale of 1-5, how would you rate your ability to complete the following tasks pertaining for the following Data Stewardship competency area:

Data Stewardship Competency Area #2: Data Policies

- *The following tasks assess your understanding on data policies.*

	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 understand and have detailed familiarity with NASA Data Policies					
I understand and have detailed familiarity with Federal Data Policies					

31. On a scale of 1-5, how would you rate your ability to complete the following tasks pertaining for the following Data Stewardship competency:

Data Stewardship Competency Area #3: Data Standards

- The following tasks assess your ability on communicating developed content on data standards.

	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 communicate the developed content for training guides to socialize stewardship related best practices					
I can lead trainings, conduct coaching sessions to upskill junior data stewards by knowledge sharing stewardship best practices					

Section 4:

Data Storage and Operations

As a Data Steward, you need to be able to maintain awareness and support data management standards relating to **data storage and operations**.

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 Storage and Operations:** 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

32. On a scale of 1-5, how would you rate your ability to complete the following tasks pertaining to data storage and operations?

	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 use my data policy awareness to adhere to NASA and federal data storage and archival policies (ex. Records Management policies)					

I can use business acumen to understand the data requirements of my respective NASA office and data domain to determine use cases for data storage and requirements to satisfy those use cases					
I can support storage solutions by partnering with business analysts and technical data custodians to create and improve storage and archival solutions					

Section 5:

Data Security

As a Data Steward, you need to have awareness of **data security** for your data domain.

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

	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 know how to interpret Federal Cybersecurity Data Policies (ex. NIST recommended security and privacy controls)					
I can implement policy and understand how cybersecurity controls apply to data sharing and accessibility					

Section 6:

Data Integration and Interoperability

As a Data Steward, you need to be able to coordinate with other data stewards and data custodians to establish data management standards relating to **data integration and interoperability** for your data domain.

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 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

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

	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 use my data domain expertise to assess data accuracy and validity					
I can partner with technical data custodians to provide requirements that improve consistency of data models across systems so data can more easily be integrated					

Section 7:

Document and Content Management

As a Data Steward, you need to be able to establish data management standards relating to **document and content management** for your data domain.

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

35. On a scale of 1-5, how would you rate your ability to complete the following tasks pertaining to document and content management?

	1: I'm not aware of	2: I'm aware but	3: I've learned but need	4: I can complete this	5: I can teach this task

	this task	need to learn	coaching	task on my own	
I can use my data domain expertise to inventory documents within a data domain					
I can control and organize documents and records through their lifecycle using process creation					

Section 8:

Reference and Data Management

As a Data Steward, you need to be able to establish data management standards relating to **reference data management** for your data domain.

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
- **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

36. On a scale of 1-5, how would you rate your ability to complete the following tasks pertaining to reference 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 use my data domain expertise in order to assess reference data accuracy and validity					
I can define reference data based on a human readable, business understanding					
I can define critical data elements for specific use cases within a data asset, and I know					

who owns the data within a data asset					
I can identify, classify, and define critical data elements					
I can create and follow processes to maintain reference data criteria and definitions in a centralized system (ex. Data Catalog, Business Glossary, Reference Data Management System)					

Section 9:

Metadata Management

As a Data Steward, you need to be able to maintain awareness and support data management standards relating to **metadata management** for your data domain.

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
- **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

37. On a scale of 1-5, how would you rate your ability to complete the following tasks pertaining to metadata 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 use my data domain expertise in order to assess metadata accuracy and validity					
I can define metadata based on a human readable, business understanding					
I can define critical data elements for specific use cases					

within a data asset, and I know who owns the data within a data asset					
I can identify, classify, and define critical data elements					
I can create and follow processes to maintain metadata criteria and definitions in a centralized system (ex. Data Catalog, Business Glossary)					

Section 10:

Golden Record Data Management

As a Data Steward, you need to be able to establish data management standards relating to **golden record data management** for your data domain.

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

38. 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 use my data domain expertise in order to understand the data and determine how to interpret and consolidate data within that domain to create golden records					
I can identify authoritative and relevant sources from which to pull information for golden records					
I can use my data governance awareness to understand who owns the authoritative data that must be attributed to a golden record					
I can assess the risks related to overmatching and undermatching when consolidating records into golden record					

Section 11:

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.

39. Please provide any additional feedback below:

Answer: _____