== What is metadata? ==

Definition: A set of data that describes and gives information about other data

There are three main types of metadata:

* Operational Metadata: Data that describes information about the operational reporting and statistics of data. This data is most often automated, but may need to be manually entered when not available and includes data related to;
  + Frequency of data load, record count, publish date, last refresh date, job execution time, job status
* Technical Metadata: Data that describes information about the structure of data and how to access the data. This data is often added into the system where data is accessed for engineering and analytics (e.g. CDP, Tableau, etc.) and includes data related to;
  + Database name, table name, view name, column name, technical data owner, security group, security classification, business group the data belongs to, upstream data source
* Business Metadata: Data that describes additional information about a piece of data that is not necessarily related to the storage or structure of data. This information is often added to a data catalog or other pre-determined source when a data catalog is not available (e.g. Excel, SharePoint, etc.) and includes data related to;
  + Business definitions, business rules, business data owner, synonyms, tags (for search)

All three types of metadata are valuable to understanding how and when to use data. Operational metadata alone tells users the status of data but does not include context of where that data exists or how it should be used. Technical metadata alone provides the data structure and classification of data, but does not allow users to determine if the data status or how it should be used. Business metadata alone describes how data should be used, but does not provide the status or structure of that data. Adding all three types of metadata to a data element will allows users to understand the complete context of the data they are using.

== Why should we manage metadata? ==

By effectively managing metadata as a data owner, users across Cargill’s data community can accurately find and reuse data rather than re-create and duplicate data in our data ecosystem. Metadata will be used to help data analysts and practitioners understand:

* Source of data
* Structure of data
* Who owns the data
* How and when the data should be used
* How to access data
* How the data fits into Cargill’s data ecosystem (see Cargill’s Data Taxonomy (link to PDT))

== Best Practices for Metadata Management ==

As Cargill implements the Data Collaboration Platform (DCP), metadata management will begin to diverge into where operational and technical metadata is stored versus where business metadata is stored. This divergence occurs so business users can add context, intended use, and data specific criteria to the data they own without needing the technical expertise to write back to or modify engineering and analytic systems that house operational and technical metadata (e.g. databases, tables, reporting systems.)

* Operational metadata is expected to reside within the system where data is accessed for engineering and analytics; automatically populated where possible or managed by technical resources
* Technical metadata is expected to reside within the system where data is accessed for engineering and analytics; populated, owned, and managed by technical resources
* Business metadata is expected to reside within DCP or another pre-determined system where data used for engineering and analytics is accessed for contextual information. In both cases the business metadata will be owned and managed in DCP by business resources following ingestion.

=== Metadata Minimums by Type===

The following metadata minimums should be met in accordance with Data Guild standards. There are a different set of minimums given the type of data that will be owned and managed. Over time, additional data sources will be mapped and updated in the tables below.

**Database Metadata**

**Table/View Metadata**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Tag | Metadata Type | Description | Example | Field in CDP | Field in DCP |
| Load Frequency | Operational | The frequency at which the data is refreshed | Daily | Load\_freq\_txt |  |
| Last Update Date | Operational | The date the data was last refreshed | 2019-07-09 12:15:54 | Last\_update\_timestamp |  |
| Table Name | Technical | Technical table name | Up\_commodity\_raw | table\_name |  |
| Table Description | Technical | A comment that describes the data in the database | Any notes about TC2 | Comment\_txt |  |
| Technical Contact Info | Technical | Publicly visible contact for any questions/comments about the data | crm-ssa-all@exch.cargill.com | Table\_owner\_id |  |
| Security Classification | Technical | How "classified" the data is according to TGRC. For more information about Cargill's Data First strategy please click [here](https://sites.cargill.com/sites/TGRC2/Pages/ssol_data_first.aspx). | RESTRICTED | Security\_class\_txt |  |
| Data Source | Technical | Technical name of source data | SFTP | Source\_txt |  |
| Security Group | Technical | The AD group for which you will need access to, to see the data | cdp\_user | Active\_directory\_read\_group\_nm |  |
| Business Definition | Business | Definition of the table/view as the business uses it | This table is used as the source of truth for customer data in CASC |  | Definition |
| Data Owner | Business | Business owner of the dataset | Rebecca Burton |  | Business Data Owner |
| Business Group | Business | Business Entity/function to which the data belongs | Commodity Contracts |  | Business Group |
| Tags | Business | A label or marker that provides some information about the thing it's attached to | Commodity |  | Tag |

**Column Metadata**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Tag | Metadata Type | Description | Example | Field in CDP | Field in DCP |
| Table Name | Technical | Technical table name | Future | table\_name |  |
| Column Name | Technical | Technical column name | Future\_id | Column\_name |  |
| Column Description | Technical | A comment that describes the data in the database | Any notes about TC2 | Comment\_txt |  |
| Business Definition | Business | Definition of the column as the business uses it | Volume is the calculation of container x load |  | Definition |
| Data Owner | Business | Business owner of the dataset | Rebecca Burton |  | Business Data Owner |
| Business Group | Business | Business Entity/function to which the data belongs | Commodity Contracts |  | Business Group |
| Tags | Business | A label or marker that provides some information about the thing it's attached to | Commodity |  | Tag |

**Additional Certified Dataset Metadata**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Tag | Metadata Type | Description | Example | Field in CDP | Field in DCP |
| Load Frequency | Operational | The frequency at which the CDS is refreshed | Daily | Load\_freq\_txt |  |
| Last Update Date | Operational | The date the data was last refreshed | 2019-07-09 12:15:54 | Last\_update\_timestamp |  |
| Source System Refresh Frequency | Technical/Business | The frequency at which the data is refreshed in upstream data sources | Daily |  |  |
| Certified Dataset Name | Technical | Technical certified dataset name | Financial\_statements\_forecast\_clean | Table\_name |  |
| Dataset Description | Technical | A description that describes the data in the dataset | Harmonized Financial Forecast Data | Comment\_txt |  |
| Certification Date | Business | Date of certification | 7/1/19 |  | Certification Date |
| Data Owner Name | Business | Dataset owner name | Rebecca Burton |  | Business Data Owner |
| Business Definition | Business | Definition of the dataset as the business refers to it | This dataset is used to measure financial forecasts for CASC |  | Definition |
| Enterprise/Group | Business | Enterprise/Function to which the data belongs | CASC |  | Enterprise/Group |
| Tags | Business | A label or marker that provides some information about the thing it's attached to | Certified Data Set |  | Tag |

**For any Metrics Created in Tables at the Column Level**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Tag | Metadata Type | Description | Example | Field in CDP | Field in DCP |
| Column Name | Technical | Technical column name | Volume | Column\_name |  |
| Column Description | Technical | A comment that describes the data that makes up the metric | The quantity of corn shipped | Comment\_txt |  |
| Enterprise/Group | Business | Enterprise/Function to which the data belongs | CASC |  | Enterprise/Group |
| Business Definition | Business | Definition of the column as the business refers to it | Volume of corn shipped |  | Description |
| Business Rule | Business | Any rule for how the column is created | Volume is the calculation of container x quantity |  | Description |
| Tags | Business | A label or marker that provides some information about the thing it's attached to | Product |  | Tag |

**Report/Dashboard Metadata**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Tag | Metadata Type | Description | Example | Field in Tableau | Field in DCP |
| Load Frequency | Operational | The frequency at which the data is refreshed in the report/dashboard | Daily | Schedule |  |
| Last Update Date | Operational | The date the data was last refreshed in the report/dashboard | 2019-07-09 12:15:54 | Last Update |  |
| Project | Operational | The name of the Project in which the report/dashboard belongs | Default | Project |  |
| Owner | Operational | The name of the owner who created the report/dashboard | Jane Doe | Owner |  |
| Business Description | Business | Description of the report/dashboard as the business uses it | This report is used to report quarterly metrics for CASC |  | Description |
| Report Owner | Business | Business owner of the report or dashboard | Rebecca Burton |  | Business Report Owner |
| Enterprise/Group | Business | Enterprise/Function to which the report belongs | CASC |  | Enterprise/Group |
| Business Group | Business | Business Entity/function to which the data belongs | Commodity Contracts |  | Business Group |
| Tags | Business | A label or marker that provides some information about the thing it's attached to | Commodity |  | Tag |

**Business Glossary and Dictionary Metadata**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Tag | Metadata Type | Description | Example | Field in DCP |
| Enterprise/Group | Business | Enterprise/Function to which the data belongs | CASC | Enterprise/Group |
| Category | Business | Business Entity/function to which the data belongs | Commodity Contracts | Category |
| Sub Category | Business | Grouping or activity within a category | Contract History | Sub Category |
| Business Term | Business | Business term name | Address fax number | Business Term |
| Technical Term | Business | Technical term name | Address\_fax\_num | Technical Term |
| Definition | Business | A comment that describes the data in the table | Contact Information: Address fax number | Description |
| Tags | Business | A label or marker that provides some information about the thing it's attached to | Commodity | Tag |
| Data Owner | Business | Business owner of the business glossary or dictionary term | Rebecca Burton | Data Owner |

==Additional Metadata==

Additional metadata may be necessary for your data to be searched and reused within your organization as well as for broader use across Cargill. Upon creating metadata, you may want to ask yourself: Do Cargill users know everything they need to know to understand this piece of data or this report? Some important information you can provide to them includes:

\* What type of analytics this data is most often used for

\* Security implications of using this data

\* Additional business glossary terms that may be helpful to understand (along these lines: Synonyms)

[[Category: Metadata]]

[[Category: Data Guild]]