Aegon Asset Management RAFT

Semantic Layer Data Governance Policy

Semantic Layer Data Governance Policy

Approval Date:	2022-07-01

Table of Contents

1. Workflow diagram 2

2. Business Data Analysts 3

2.1. What is the standard? 3

2.2. How do they do it? 3

3. Data Officer 4

3.1. What is the standard? 4

3.2. How do they do it? 4

4. Data Owner 4

4.1. What is the standard? 5

4.2. How do they do it? 5

5. Data Officer .... **Error! Bookmark not defined.**

5.1. What is the standard? .... **Error! Bookmark not defined.**

5.2. How do they do it? .... **Error! Bookmark not defined.**

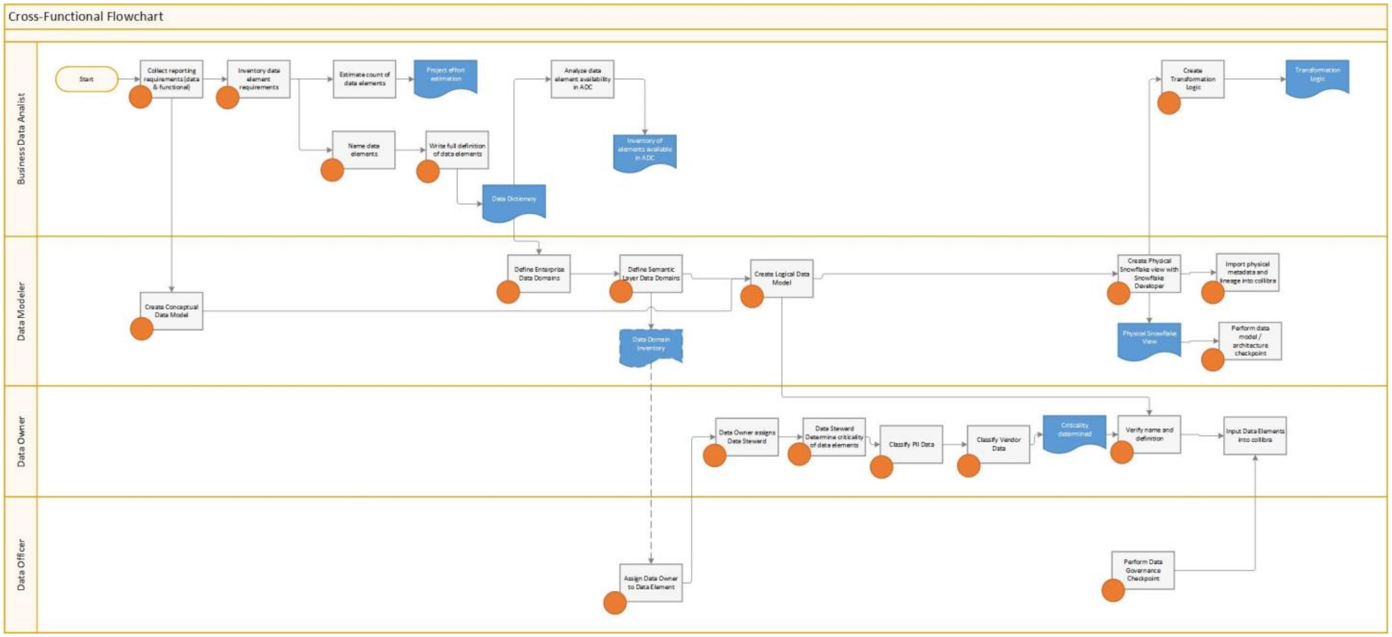
6. Appendix 7

6.1. Requestor spreasheet (best practice) 7

6.2. PII classification spreadsheet (best practice) 8

6.3. Data element criticality 9

1. Workflow diagram


Cross-Functional Flowchart
Business Data Analist
Data Modeler
Data Owner
Data Officer
Start
Collect reporting requirements (data & functional)
Inventory data element requirements
Estimate count of data elements
Project effort estimation
Name data elements
Write full definition of data elements
Data Dictionary
Analyze data element availability in ADC
Inventory of elements available in ADC
Create Transformation Logic
Transformation logic
Create Conceptual Data Model
Define Enterprise Data Domains
Define Semantic Layer Data Domains
Data Domain Inventory
Create Logical Data Model
Create Physical Snowflake view with Snowflake Developer
Import physical metadata and lineage into collibra
Physical Snowflake View
Perform data model / architecture checkpoint
Data Owner assigns Data Steward
Data Steward Determine criticality of data elements
Classify PII Data
Classify Vendor Data
Criticality determined
Verify name and definition
Input Data Elements into collibra
Assign Data Owner to Data Element
Perform Data Governance Checkpoint

2. Business Data Analysts

Collect reporting requirements: As a best practice, reporting requirements can be collected by filling out the Requestor Spreadsheet which can be found in the appendix.

Name data elements: The act of naming and the data element to the best of your ability.

Write full definitions: The act of clearly describing the concept to the best of your ability.

2.1. What is the standard?

2.1.1. Reporting Requirement Standards

1. Reporting requirements are data driven consumption needs from business teams
2. Reporting requirements are collected in a structured manner.
3. Reporting requirement are documented and maintained in a single source of truth document (e.g. Requestor spreadsheet)

2.1.2. Naming Standards

1. Every data element is assigned a name
2. The name of the data element must stay the same across work products and systems
3. The name of data elements are documented and maintained in the data dictionary which serves as the single source of truth for data element names.
4. Names must be clear, descriptive and follow a consistent format

2.1.3. Definition Standards

1. Every data element is assigned a description
2. The description of the data element must stay the same across work products and systems
3. The description of data elements are documented and maintained in the data dictionary which serves as the single source of truth for data element descriptions.
4. Business teams own their definitions and are responsible for keeping them up to date
5. Definitions should be unique and distinguishable from other definition

2.2. How do they do it?

2.2.1. Reporting Requirement Procedure

1. As a best practice, reporting requirements can be collected by filling out the Requestor Spreadsheet which can be found in the Appendix.

2.2.2. Naming Best Practices

2. Data element names are always built using a prime word, a qualifier, and a class word.
3. The data element name always begins with a prime word.
4. The data element name always ends with a class word (except for surrogate keys in subtype entities as described further in this document).
5. Use qualifiers to enhance the prime word and further describe the element name.
6. Data element names are composed as follows: **Prime (P) / qualifier (Q) / class (C) =**

(P)________________/(Q)_______________/(C)_____________________

7. The prime word is a business entity name that has a lasting importance to the business or enterprise (*e.g. people, places, things*).
8. A qualifier is a word or words that provide further business meaning to the prime word, like an adjective. A qualifier is not required in a element name. (*e.g. order, category rank*).

9. A class word is a type of business term that indicates the nature of the data represented by the element. The class word is the method of categorising the element to a specific type of data (*e.g. image, number, percent, identifier*).

2.2.3. Definitions Best Practices

1. The definition should use words that have a precise meaning. Definitions should not contain words that have multiple meanings or multiple word senses.
2. The definition should use the shortest description possible that is still clear.
3. The definition should not use the term being defined in the definition itself. This is known as a circular definition.
4. The definition should not be an inverse description, i.e. a description of what it is not.
5. The definition should differentiate a data element from other data elements, i.e. should not be applicable to multiple data elements. This process is called disambiguation.
6. Definitions should not contain acronyms, abbreviations or initials.
7. Definitions should not refer to terms or concepts that might be misinterpreted or that have different meanings based on the context of a situation.
8. Definitions should not be a reordering of the words in the data element name.

3. Data Officer

Assign Data Owners: The act of assigning and approving a Data Owner role to a Data Element.

3.1. What is the standard?

5.1.1. Assign Data Owners

1. Every Data Element must be assigned to a Data Owner.
2. A Data Owner is accountable for the care of their data elements.

3.2. How do they do it?

5.2.1. Assign Data Owners

1. Select a data owner that is familiar with the theme/data domain of selected data elements (*e.g. financial instrument or customer)*.
2. Ensure the candidate data owner is familiar with Data Ownership standards.
3. Ensure the candidate data owner is familiar with Data Stewardship standards.
4. Ensure the candidate data owner is a subject matter expert and is familiar with the business usage of the data elements at hand.

4. Data Owner

Assigning Data Steward: The act of assigning and approving a Data Steward to a Data Element.

Determine Criticality: The act of assessing the importance of a Data Element to key business processes by taking into account the ramifications and risks to the organization if the Data Element (and related physical fields) is incorrect or unavailable.

Classify PII: The act of discovering and classifying privacy characteristics of data elements.

Classify Vendor: The act of discovering and classifying data source characteristics of data elements.

Verify naming and definitions: The act of reviewing the proposed name and definition of data elements.

4.1. What is the standard?

4.1.1. Assign stewards

1. The Data Owner assigns the Data Steward(s) to all of their data elements.
2. A Data Steward is responsible for the care of their data elements.

4.1.2. Criticality

1. Data profiling, data lineage and data quality efforts are prioritized based on criticality.
2. Criticality is documented and maintained in the data dictionary.

4.1.3. Classify PII

1. PII data is subject to adequate safety measures in order to ensure regulatory compliance and prevent harm to customer or company.
2. PII classification is documented and maintained in the data dictionary.

4.1.4. Classify Vendor Data

1. Different internal and external data sources are documented and maintained in the data dictionary.

4.1.5. Verify name and definition

1. The compliance of names and definitions with the naming and definition standards is ensured.
2. The consistency of names and definitions across work artificats (e.g. *data models and data dictionaries*) is ensured.

4.2. How do they do it?

4.2.1. Assign data steward

1. Familirarity with the data ownership standards.
2. Familirarity with data stewardship standards.
3. Familiarity with the technical requirements of the data element.
4. The Data Steward approves the maintenance process.
5. The Data Steward assigns the agreed maintenance process to the data element.

4.2.2. Determine criticality

1. Understand the impact of data elements on regulatory reporting, operational performance and business intelligence.
2. Differentiate between data elements that are required and not requried for regulatory reporting, operational performance and business intelligence.
3. Assign a criticality level of 0-3 as demonstrated in Appendix 5.3.
4. Data is considered critical when it is directly or indirectly used for:
    a. Supervisory, statistical and tax reporting
    b. Financial reporting and other public disclosures
    c. Management reporting and commercial decision making
    d. Customer management

4.2.3. Classification PII data

1. Understand the subject matter of the data element.
2. Differentiate between PII data and non-PII data using the PII classification spreadsheet.
3. Assign the PII level of ‘non-PII’, ‘PII-Analytical’ ‘PII-Operational’ to all data elements in the data dictionary.

4.2.4. Classification Vendor data

1. Understand the ultimate source of the data element.
2. Differentiate between internal and external data elements.
3. Differentiate between different internal data sources (*e.g. Business Units, regions*).
4. Differentiate between different external data sources (*e.g. Bloomberg, open sources*).
5. Assign the internal or external source name to the data element in the data dictionary.

4.2.5. Verify name and definition

1. Understand the subject matter of the data element.
2. Familirairty with the naming and definition standards and procedures.
3. Check whether each name meets the naming and definition standards and procedures.
4. Correct names that do not meet the standards and procedures.

5. Appendix

5.1. Requestor spreasheet (best practice)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
REF #	REQUESTED TARGET DBMS (Snowflake, Oracle, etc.)	SOURCE SYSTEM / DATA SET / COPYBOOK	SOURCE TABLE NAME	SOURCE FIELD NAME	SOURCE DATA TYPE	SOURCE LENGTH	SOURCE DECIMAL	SOURCE NULLS (Y/N)	TARGET FIELD NAME	TARGET DATA TYPE	TARGET LENGTH	TARGET DECIMAL	TARGET NULLS (Y/N)	REQUESTER BUSINESS DEFINITION	CODE VALUES	CALCULATIONS	IN CURRENT DATABASE ?(Y/N)	HISTORY REQUIRED (Y/N)	HISTORY CAPTURE TYPE (SNAPSHOT/ CHANGE)	HISTORY CAPTURE FREQUENCY (RATE OF CHANGE)

Commented [A1]: Add a few sample rows here

1. In the column, specify the data base management system this data base will reside in upon completion.
2. Enter the name of the system, data base, copybook, etc., this field/column originates from, if any.
3. Indicate the name of the table where the field/column resides, in the data base of origin, if any.
4. Enter the name of the field/column in the data base of origin, if any.
5. Enter the name of the field/column in the data base of origin, if any.
6. Indicate the number of bytes this field/column occupied in the data base of origin, if any.
7. Indicate the number of spaces to the right of the decimal point, this field/column occupied in the data base of origin, if any.
8. You may enter whether this field/column was optional or mandatory in the former or existing data base of origin, if any
9. If your business has a 'preferred' name it would like this field/column to be called in the new data base, enter it here for consideration.
10. In this column, enter the data type this field/column will need to be in the new data base. Ex.: char, integer,date, decimal, time, smallint, varchar, varchar2
11. In this column, enter the number of bytes you think you will need for this field/column in your new data base.
12. In this column, enter the number of decimal spaces to the right of the decimal point that you think you need. Ex.: 2.
13. In this column, you may indicate if this field/column is optional or mandatory.
14. Please enter a meaningful definition that explains exactly what this field/column means. (The repository name will be based on this.)
15. In this column, indicate any codes and the values for the codes. Example of format: C = Code
16. In this column, enter any calculations used to derive information that will populate the field/column.
17. This column should be used only if this is a request for additions or changes to an existing database.
18. This column is only for BI/Analytics projects. Complete to indicate whether History Capture is required for this element.
19. This column is only for BI/Analytics projects. Complete to indicate how often History is Captured.
20. This column is only for BI/Analytics projects. Complete to indicate how often the data is expected to change.

5.2. PII classification spreadsheet (best practice)

Data Class Category	Class Code	Class Type	Class Definition
Non-PII	NP	Non-PI / Non-PII	Data element does not contain any Personal Information (PI) or Person Identifiable Information (PII)
PII-Analytical	A1	Analytic Data - Personal	Data elements within analytic class can be used for analysis and trending; however, must not be sufficient to distinguish particular individuals directly even when aggregated or joined with other elements from this class. A1 is associated with personal details E.g. Legal Status, Language Preference etc
	A2	Analytic Data - Financial	A2 is associated with financial details. E.g. Income, debt, Asset, net worth. Financial trans, balances etc
	A3	Analytic Data – Secure Personal	A3 is associated with secure personal information, requiring additional approvals E.g. Ethnicity, Gender etc
	A4	Analytic Data – Secure Financial	A4 is associated with secure financial information, requiring additional approvals E.g. Credit Score
PII-Operational	O1	Contact Data	Data elements within this class can be used to locate or communicate with an individual. The data elements are sufficient to distinguish an individual by themselves. E.g. name, physical address, e-mail address.
	O2	Internal Identifier	Data Elements within this class can be used to identify a customer within an internal operational processes or between entities. E.g. employee ID number, student ID number, credit bureau personal identification number.
	O3	Authentication	Data Elements within this class can be used for authentication of a customer within operational processes. E.g. password, mother's maiden name, PIN, security Q&A etc
	O4	Account Identifier / device	Data elements within this class are required to perform an operational financial transaction. E.g. Savings, Money Market account number.
	O5	Customer Identifier - Govt. Primary	Data Elements within this class can be used to identify a customer within operational processes that requires government unique identifier. E.g. US - SSN for tax reporting, UK – National Insurance Number
	O6	Customer Identifier - Govt. Secondary	Data Elements within this class can be used to identify a customer within operational processes. E.g. driver's license # for auto loan, passport #, Medicare # and associated distinguishable attributes such as issue city, issue date, expiry date etc
	O7	Unstructured sensitive data	PI elements within this grouping contain a combination of PI data elements and often reside within a specific document such as a loan application (e.g. name, SSN, income, race, gender, credit score, etc.). Free-form fields in the database could contain potential PII information and hence categorized into this class of data

5.3. Data element criticality

Criticality Scale	Criticality Name	Description	Regulatory /Compliance	Effect
Level 3	Extremely Critical	Data and metadata that is critical to staying open for business and protected by a business continuity plan that would allow for resumption of services immediately.	Vital regulatory/compliance field that must have an accrued history of values for proper reporting.	Automatic urgent priority for data incidents - 3 hour resolution time
Level 2	Critical	Data and metadata required to administer functions within the business, but not immediately required by a business continuity plan.	Possibly related to regulatory/compliance field.	Automatic high priority for data incidents - 8 hour resolution time
Level 1	Not Critical	Data and metadata that is not critical to remain open for business. This data and metadata does not play a significant role in daily operations, but may be used for reporting.	No relation to regulatory/compliance fields.	Low/Medium rating for data incidents - 3-5 day resolution time
Level 0	Potentially Obsolete	Data and metadata that is no longer used and is candidate for removal.	No relation to regulatory/compliance fields.	In case no usage is found after completion of data lineage, the field will be removed from systems.