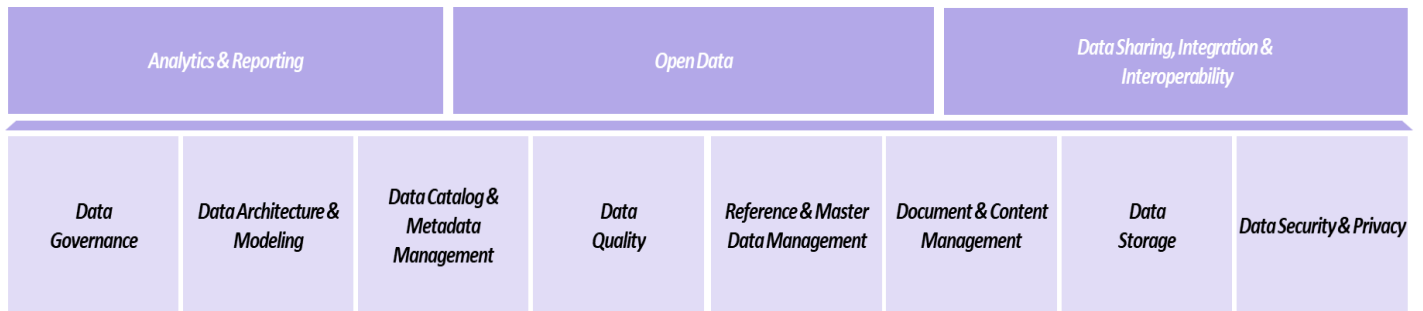




Data Catalog Implementation Guidelines

Overview

The implementation guidelines were developed based on the controls governing the implementation of the 11 data management domains:



This document describes more specifically the data catalog implementation guidelines, the processes that enable data cataloging and the roles and responsibilities that enable its development and maintenance.

The data catalog shall serve as a repository to collect information about an ADGE's data in a single location where this information can be accessed and maintained more conveniently. Furthermore, though some data may be relatively well understood, it is unlikely that even this data is described as comprehensively as it should be.

The data catalog consists of three artifacts:

- Dataset Inventory: document that incorporates a full list of the ADGE's data
- Business Glossary: repository of metadata that adds semantic context to data and defines the business terms and definitions related to physical assets so that users can easily collaborate and understand the semantics behind data and how it is being used
- Data Dictionary: repository that includes detailed definition and description of datasets and their attributes (business and technical metadata)

As part of these data catalog implementation guidelines, 4 processes are addressed:

1. Create/Update Dataset Inventory
2. Populate Business Glossary and Data Dictionary
3. Publish Data Catalog
4. Use Data Catalog

All data cataloging processes currently leverage templates provided by ADDA that are shared back and forth between the ADGE and ADDA and within the ADGE itself. At a later stage, all these activities will be done through the data catalog platform which will be provisioned by ADDA and for which further information will be provided.

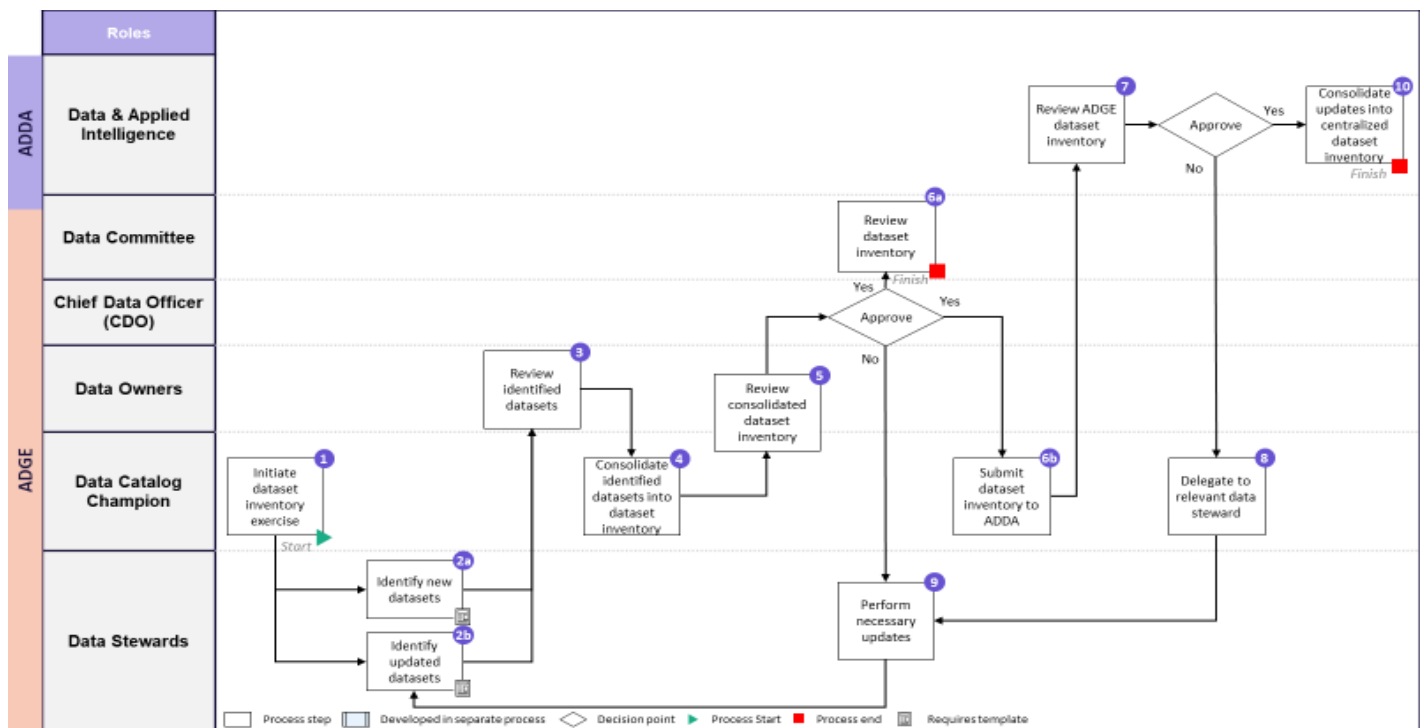
Create/Update Dataset Inventory

Overview

Each ADGE needs to develop a dataset inventory that consists of the list of all its data. This will enable the ADGE to have an overview of the data it possesses, and consequently prioritize this data and identify which datasets should have their business and technical metadata populated first.

This process should be triggered when an ADGE develops its dataset inventory for the first time and then becomes a continuous process through which the dataset inventory is modified or updated.

Process



Activities

1 – Initiate dataset inventory exercise – Data Catalog Champion

- Trigger dataset inventory implementation exercise as part of the data catalog implementation plan
- Meet with data stewards and relevant stakeholders and assign roles and responsibilities

2a – Identify new datasets– Data Stewards

- Identify new datasets which are not part of the dataset inventory yet
- For each identified dataset populate the following metadata description within the dataset inventory template currently adopted by ADDA (refer to the Templates section for an extract of the dataset inventory template):
 - Dataset ID
 - Title (if it doesn't have an existing title that you are aware of, choose a short, descriptive name) (both in Arabic and English)
 - Description: clarify what data is being referred to and its scope (both in Arabic and English)

- Owner: the department responsible for the data
- Source(s)
- Size: number of records in this dataset
- Update frequency (daily, weekly, monthly, yearly)
- Date added
- Notes
- At a later stage, the templates will be replaced by the data catalog platform which will be provisioned by ADDA and for which further information will be provided

2b – Identify updated datasets– Data Stewards

- Identify datasets that includes information that may need to be updated in the dataset inventory
- If needed, update the values listed in the dataset inventory for the identified datasets

3 – Review identified datasets– Data Owners

- Review each identified dataset and ensure they have been added appropriately without missing any dataset
- Share the list of identified datasets with the Data Catalog Champion for it to be consolidated into the ADGE’s centralized dataset inventory

4 – Consolidate identified datasets into dataset inventory – Data Catalog Champion

- Consolidate dataset lists received from all Data Owners into a single list to create the ADGE’s centralized dataset inventory

5 – Review consolidated dataset inventory – Data Owners and Data Catalog Champion

- Review centralized dataset inventory, identify discrepancies such as duplicated datasets and take mitigating measures to minimize them
- In case of dataset duplication for example, debate and agree on the one that should be used or escalate to ADGEs Data Committee if the issues couldn’t be resolved on the Data Owners level

6a – Review dataset inventory – Data Committee

- Review dataset inventory and ensure ADGE strategic priorities and objectives are met and ADDA’s established targets are respected
- Identify potential issues and communicate them with the Data Catalog Champion
- Resolve any issues that couldn’t be resolved by data owners
- This step is a reporting step to keep the Data Committee aware on the progress and status of the dataset inventory and should not be considered as an approval step

6b – Submit dataset inventory to ADDA – Data Catalog Champion

- Share dataset inventory with ADDA for review and feedback

7 – Review ADGE dataset inventory– Data & Applied Intelligence

- Review ADGE dataset inventory, validate that it fulfills the dataset inventory guidelines, standards and policies provided by ADDA in the Abu Dhabi Government Data Management Standards and Policies
- Identify issues or areas of improvement and communicate feedback to ADGE

8 – Delegate to relevant data steward– Data Catalog Champion

- Receive issue detected by ADDA
- Perform a root cause analysis on the dataset in question
- Delegate to relevant data steward to fix the issue

9 – Perform necessary updates – Data Stewards

- Based on the Data Catalog Champion’s assessment, take necessary actions to update impacted dataset
- Once the update or fix is done, revert to Data Catalog Champion with the solution details

10 – Consolidate updates into centralized dataset inventory – Data & Applied Intelligence

- Incorporate the ADGE dataset inventory into the centralized Abu Dhabi dataset inventory owned and managed by ADDA
- Notify ADGE that its dataset inventory has been incorporated in the centralized Abu Dhabi dataset inventory

Templates

Step 2a-2b: Dataset Inventory Template Extract

Dataset List سجل مجموعات البيانات										
Dataset ID رقم مجموعة البيانات	Title Ar إسم مجموعة البيانات	Title En	Description Ar وصف مجموعة البيانات	Description En	Owner مالك مجموعة البيانات	Source(s) المصدر	Size الحجم (number of rows)	Update Frequency فترة التحديث (daily, weekly, monthly, yearly)	Date Added تاريخ الإضافة	Notes ملاحظات
DS-										
DS-										
DS-										

Governance

Roles \ Activities	Roles					
	Data Committee	Chief Data Officer	Data Owner	Open Data Champion	Data Steward	Data and Applied Intelligence
1 - Initiate dataset inventory exercise		A	C/I	R	I	C
2a – Identify new datasets			A	C/I	R	C
2b – Identify updated datasets			A	C/I	R	C
3 – Review identified datasets		A	R	C	I	
4 – Consolidate identified datasets into dataset inventory		A	I	R	I	
5 - Review consolidated dataset inventory		A	R	C		
6a - Review dataset inventory	A/R	C		I		

6b - Submit dataset inventory to ADDA		A	C/I	R		
7 - Review ADGE dataset inventory				I		A/R
8 – Delegate to relevant data steward		A	C	R	I	
9 - Perform necessary updates		I	A	C	R	C
10 - Consolidate updates into centralized dataset inventory				I		A/R

R: Responsible – *Person who does the task*

A: Accountable – *Person who delegates work and provides final review on a task or deliverable before it is deemed complete*

C: Consulted – *Person who provides input on a task or deliverable based on the impact on their work or their domain of expertise*

I: Informed – *Person who needs to be kept in the loop on task or deliverable progress*

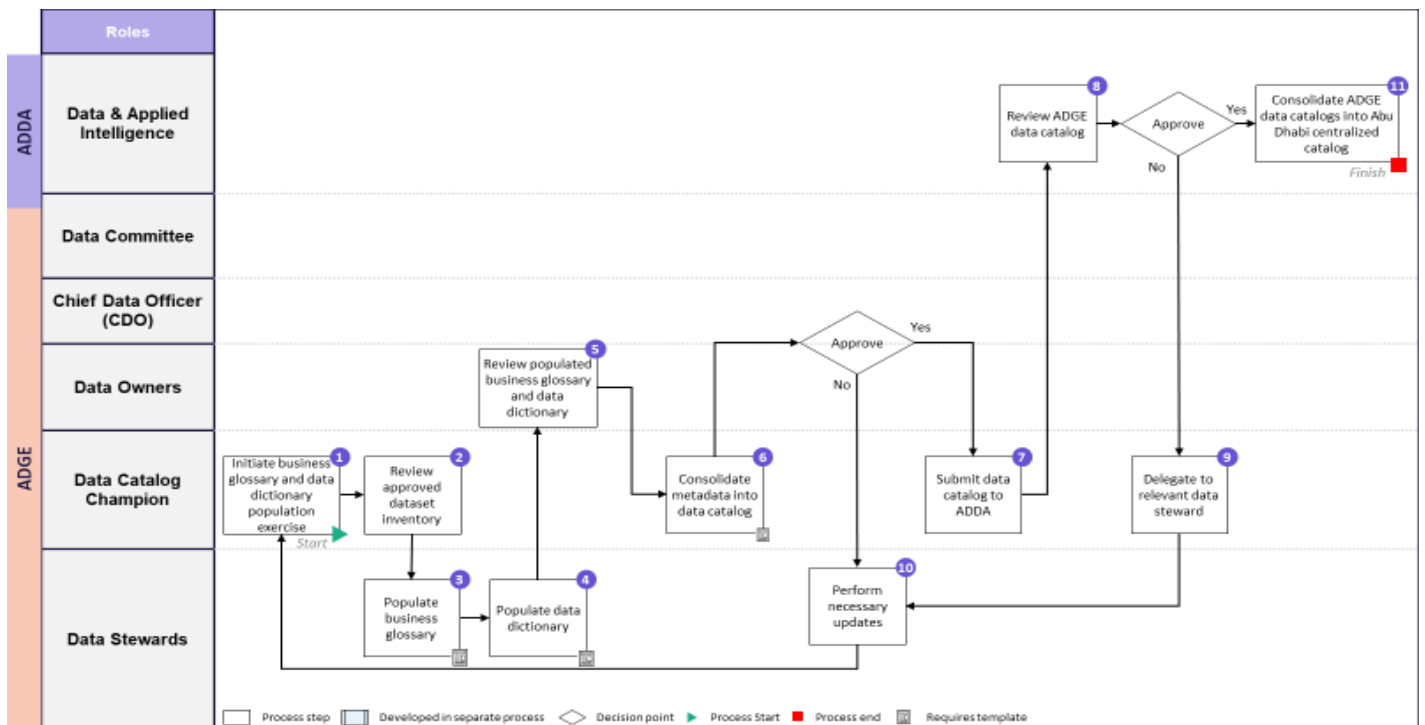
Populate Business Glossary and Data Dictionary

Overview

The business glossary and data dictionary are simply ways to describe data, using metadata (data about data). A data catalog is a repository of metadata that combines both the business glossary and data dictionary. Therefore, the Business Glossary and Data Dictionary process is an essential step in the process of creating or updating a data catalog because it describes how an ADGE can identify and populate metadata for the datasets identified in its dataset inventory.

This process should be triggered when an ADGE develops its data catalog for the first time and then becomes a continuous process through which the business glossary, data dictionary and any relevant metadata need is modified or updated.

Process



Activities

1 – Initiate business glossary and data dictionary population exercise – **Data Catalog Champion**

- Trigger business glossary and data dictionary population exercise as part of the data catalog implementation plan
- Meet with data stewards and relevant stakeholders and assign roles and responsibilities

2 – Review dataset inventory – **Data Catalog Champion**

- Review dataset inventory and check for datasets for which business glossary and data dictionary were not populated. These will be the datasets for which the business glossary and data dictionary will be populated as part of this process
- The first business glossary and data dictionary population exercise will need to be done on all datasets identified as part of the dataset inventory exercise

3 – Populate business glossary – Data Stewards

- Data stewards should populate the business glossary based on the template provided by ADDA (refer to the Templates section for an extract of business glossary with some suggested attributes). A business glossary is a repository of metadata that adds semantic context to data. It enables an organization to build trust and create accountability through common business terminologies. It defines the business terms and definitions so that employees and relevant stakeholders can easily collaborate, understand the semantics behind data and how it is being used, including usage of business and data definitions, taxonomies, critical data attributes, KPIs, metrics, etc. This vocabulary furnishes clear meaning and business context and is linked to the underlying metadata to provide an association between terms and objects. Business metadata which constitutes a business glossary consists of the content and condition of data and includes non-technical names and definitions
- At a later stage, the templates will be replaced by the data catalog platform which will be provisioned by ADDA and for which further information will be provided

4 – Populate data dictionary – Data Stewards

- Populate metadata for each of the datasets defined in dataset inventory in the data dictionary based on the template provided by ADDA (refer to the Templates section for an extract of the data dictionary with some suggested attributes). A data dictionary is a detailed definition and description of datasets and their fields. It contains business and technical metadata that describes information such as data element name, description, type, format, length, allowed values, default values, constraints, relations to other data elements, etc.
- Business metadata will be filled by data stewards while technical metadata will be filled by the ADGE's IT team
- At a later stage, the templates will be replaced by the data catalog platform which will be provisioned by ADDA and for which further information will be provided

5 – Review populated business glossary and data dictionary – Data Owners

- Review both the business glossary and the data dictionary and ensure that the metadata has been correctly populated for each dataset and that no metadata field is missing
- Share business glossary and data dictionary with Data Catalog Champion for consolidation into the ADGE's centralized data catalog

6 – Consolidate metadata into data catalog – Data Catalog Champion

- Consolidate business glossaries and data dictionaries received from all Data Owners to create the ADGE's centralized data catalog
- Refer to the Templates section for an extract of a data catalog with some suggested attributes

7 – Submit data catalog to ADDA – Data Catalog Champion

- Share data catalog with ADDA for review and feedback

8 – Review ADGE data catalog – Data & Applied Intelligence

- Review ADGE data catalog, validate that it fulfills all the guidelines, standards and policies provided by ADDA the Abu Dhabi Government Data Management Standards and Policies
- Identify issues or areas of improvement and communicate feedback to ADGE

9 – Delegate to relevant data steward – Data Catalog Champion

- Receive issue detected by ADDA
- Perform a root cause analysis on the dataset or metadata in question
- Delegate to relevant data steward to fix the issue

Governance

<div style="text-align: right;">Roles</div> <div style="text-align: left;">Activities</div>	Data Committee	Chief Data Officer	Data Owner	Open Data Champion	Data Steward	Data and Applied Intelligence
1 - Initiate business glossary and data dictionary population exercise		A	C/I	R	I	C
2 – Review dataset inventory		A	C/I	R	I	
3 – Populate business glossary			A	C/I	R	
4 - Populate data dictionary			A	C/I	R	
5 – Review populated business glossary and data dictionary		A	R	C	I	
6 – Consolidate metadata into data catalog		A	I	R	I	
7 - Submit data catalog to ADDA		A	I	R		
8 - Review ADGE data catalog				I		A/R
9 – Delegate to relevant data steward		A	C	R	I	
10 - Perform necessary updates		I	A	C	R	C
11 - Consolidate ADGE data catalogs into Abu Dhabi centralized catalog				I		A/R

R: Responsible – *Person who does the task*

A: Accountable – *Person who delegates work and provides final review on a task or deliverable before it is deemed complete*

C: Consulted – *Person who provides input on a task or deliverable based on the impact on their work or their domain of expertise*

I: Informed – *Person who needs to be kept in the loop on task or deliverable progress*

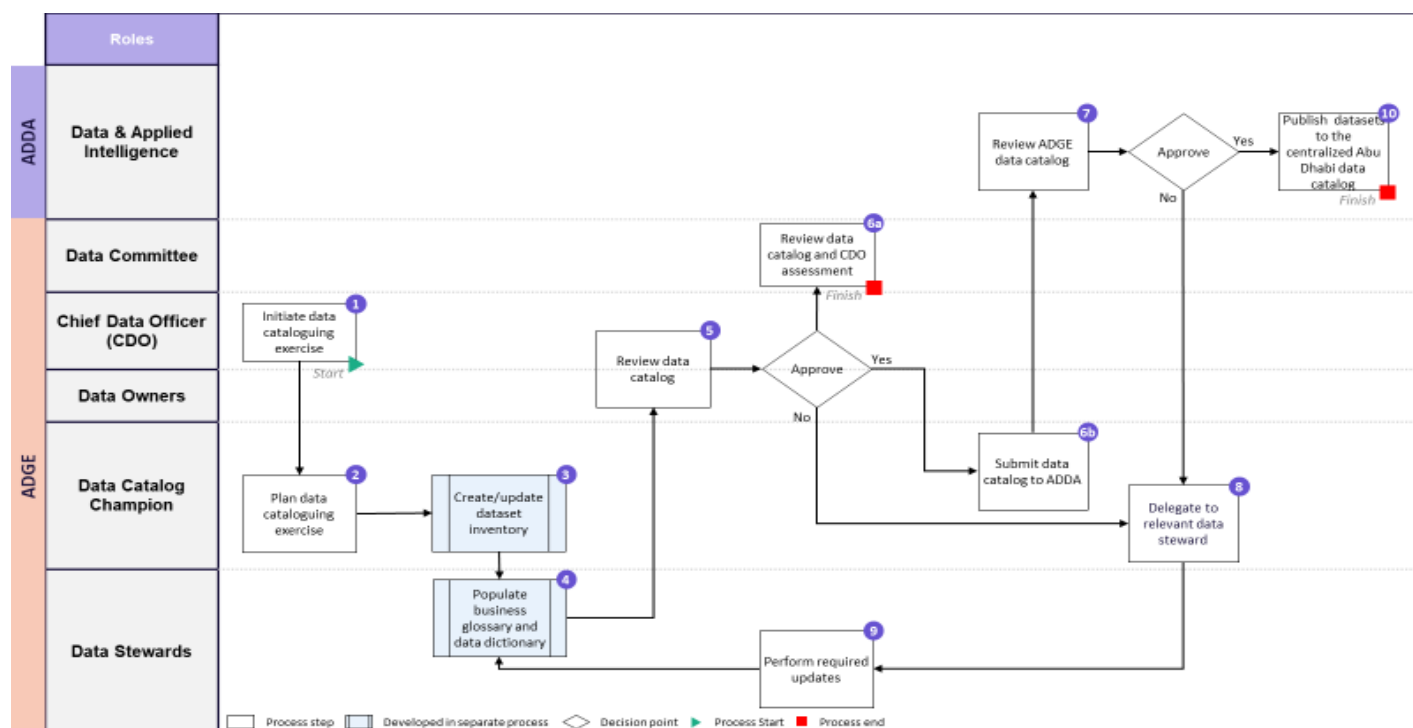
Publish Data Catalog

Overview

The process identified in this section serves the purpose of helping ADGEs develop, publish (make accessible for authorized users) and enhance their data catalog. This process connects to some of the other processes in this guideline.

This process should be triggered when an ADGE develops its data catalog for the first time and then becomes a continuous process through which data catalog is modified or updated.

Process



Activities

1 – Initiate data cataloguing exercise – Chief Data Officer (CDO)

- Activate operating model roles related to the data catalog
- Adopt standards and policies defined by ADDA in the Abu Dhabi Government Data Management Standards and Policies
- Trigger data catalog planning and implementation exercise

2 – Plan data cataloguing exercise – Data Catalog Champion

- Meet with data stewards and relevant stakeholders and assign roles and responsibilities
- Develop plan for initial data catalog implementation and data catalog maintenance

3 – Create/update dataset inventory – Data Catalog Champion

- Each ADGE needs to develop a dataset inventory that includes a list of all its datasets. This will enable the ADGE to have an overview of the data it holds, and consequently prioritize its datasets and identify which datasets should have their business and technical metadata populated first
- Refer to the Create/Update Dataset Inventory section of this document for the detailed activities

4 – Populate business glossary and data dictionary – Data Stewards

- The business glossary and data dictionary provide valuable context and meaning to data that will dramatically increase the usability and discoverability of the data. Each dataset should be described with business and technical metadata some of which is mandatory for all datasets and others optional based on the nature of the dataset
- Refer to the Populate Business Glossary and Data Dictionary section of this document for the detailed activities

5 – Review data catalog – Chief Data Officer and Data Owners

- Review data catalog and approve for sharing with ADDA
- In case an issue is detected, revert to the Data Catalog Champion for updates
- Report and share the reviewed catalog with the ADGE Data Committee for their discussion, assessment and revision

6a – Review data catalog and CDO assessment – Data Committee

- Review data catalog and ensure ADGE strategic priorities and objectives are met and ADDA's established targets are respected
- Identify and resolve potential issues or risks and communicate them with the CDO and Data Catalog Champion
- This step is a reporting step to keep the Data Committee aware on the progress and status of the data catalog and should not be considered as an approval step

6b – Submit data catalog to ADDA – Data Catalog Champion

- Share data catalog with ADDA for review and feedback

7 – Review ADGE data catalog – Data & Applied Intelligence

- Receive and review data catalog shared by ADGE and approve for consolidation in the Abu Dhabi centralized data catalog
- Ensure that all entries abide by the data cataloging guidelines, standards and policies provided by ADDA in the Abu Dhabi Government Data Management Standards and Policies
- In case of any issues with certain catalog entries are detected, revert to the relevant ADGE with an explanation of the issue

8 – Delegate to relevant data steward – Data Catalog Champion

- Receive issues detected by ADDA from CDO
- Perform a root cause analysis on the data catalog entry in question
- Delegate to relevant data steward to fix the issue

9 – Perform required updates – Data Stewards

- Based on the Data Catalog Champion's assessment, take necessary actions to update impacted catalog entries
- Once the update or fix is done, revert to Data Catalog Champion with the solution details

10 – Publish new datasets to the centralized Abu Dhabi data catalog – Data & Applied Intelligence

- Incorporate the ADGE data catalog into the centralized Abu Dhabi data catalog owned and managed by ADDA

- Notify ADGE that their data catalog has been incorporated in the centralized Abu Dhabi data catalog

Templates

No templates were identified as part of this process

Governance

<div style="text-align: center;">Roles</div> <div style="text-align: center;">Activities</div>	Data Committee	Chief Data Officer	Data Owner	Open Data Champion	Data Steward	Data and Applied Intelligence
1 – Initiate data cataloging exercise		A/R	C	C	I	I
2 – Plan data cataloging exercise		A	C/I	R	I	C
3 – Create/update inventory		A	C	R	R	C/I
4 – Populate business glossary and data dictionary			A	C/I	R	C/I
5 – Review data catalog		A/R	R	C	C	
6a – Review data catalog and CDO assessment	A/R	C		I		
6b – Submit data catalog to ADDA		A	C/I	R		
7 – Review ADGE data catalog				I		A/R
8 – Delegate to relevant data steward		A	C	R	I	
9 – Perform required updates		I	A	C	R	C
10 - Publish new datasets to the centralized Abu Dhabi data catalog				I		A/R

 Process

R: Responsible – *Person who does the task*

A: Accountable – *Person who delegates work and provides final review on a task or deliverable before it is deemed complete*

C: Consulted – *Person who provides input on a task or deliverable based on the impact on their work or their domain of expertise*

I: Informed – *Person who needs to be kept in the loop on task or deliverable progress*

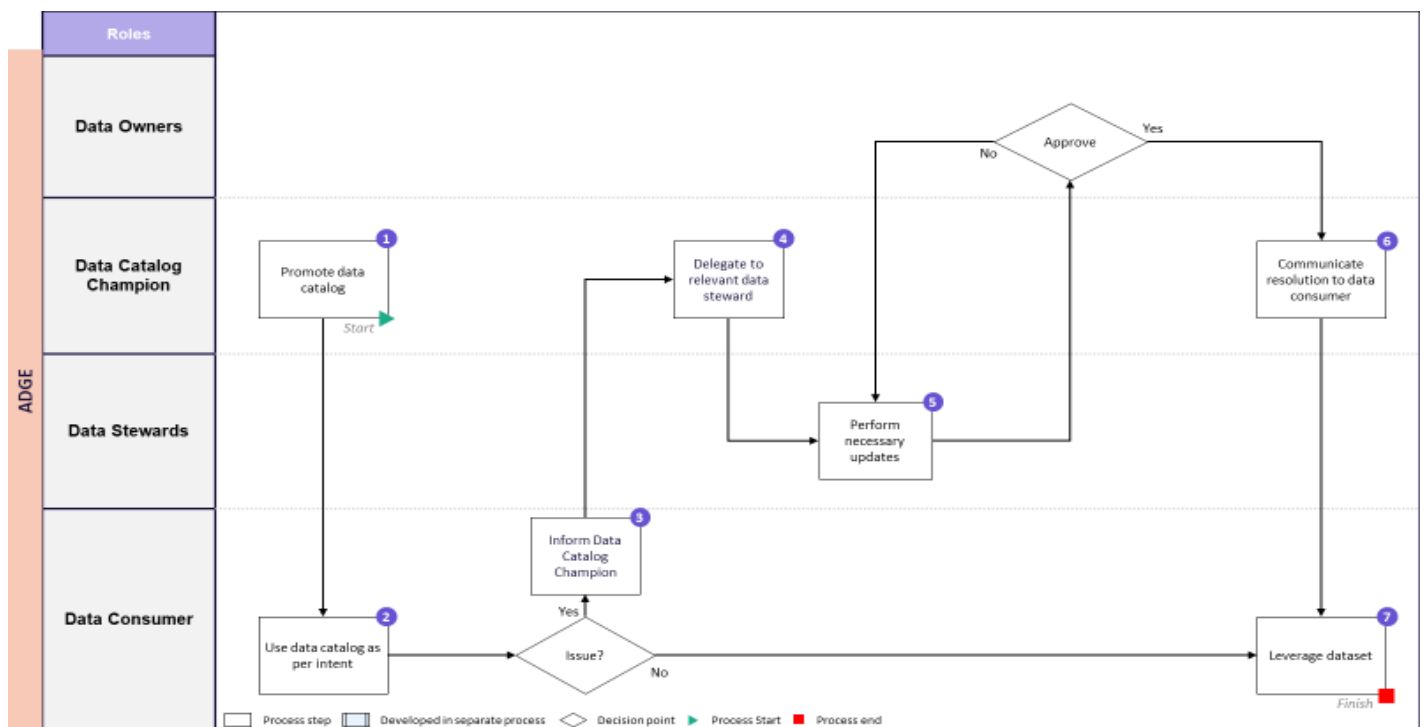
Data Catalog Usage

Overview

The objective of this process is to guide users on how to leverage their ADGE's data catalog and enable them to search for data assets to satisfy their information needs.

This process is a continuous one, it is triggered when a data consumer uses the ADGE data catalog or detects an issue in it.

Process



Activities

1 – Promote data catalog – Data Catalog Champion

- Develop and send ADGE wide communication announcing the creation of the ADGE data catalog
- Promote the data catalog functionalities through trainings and tutorials
- Create a continuous communication channel dedicated to the data catalog that is accessible by the ADGE employees to allow them to share ideas thus enabling collaboration and continuous improvement

2 – Use data catalog as per intent – Data Consumer

- Use the data catalog to search for and discover data assets that can address the specific purpose of the request

3 – Inform Data Catalog Champion – Data Consumer

- In case an issue is detected while browsing through the data catalog or more specifically in the entry the user would like to use, contact the Data Catalog Champion to inform him and provide him details about the issue

4 – Delegate to relevant data steward – Data Catalog Champion

- Receive issue detected by a data consumer
- Perform a root cause analysis on the dataset or metadata in question
- Delegate to relevant data steward to fix the issue

5 – Perform necessary updates – Data Stewards

- Based on the Data Catalog Champion’s assessment, take necessary actions to update impacted dataset or metadata
- Once the update or fix is done, revert to Data Catalog Champion with the solution details

6 – Communicate resolution to data consumer – Data Stewards

- Send notification to data consumers about new changes in the data catalog

7 – Leverage dataset – Data Consumer

- Access or request access to the dataset searched for in the data catalog

Templates

No templates were identified as part of this process

Governance

Roles Activities	Data Committee	Chief Data Officer	Data Owner	Open Data Champion	Data Steward	Data Consumer	Data and Applied Intelligence
1 - Promote data catalog		A		R		I	
2 – Use data catalog as per intent						A/R	
3 – Inform Data Catalog Champion				I		A/R	
4 - Delegate to relevant data steward		A	C	R	I		
5 – Perform required updates		I	A	C	R		C
6 – Communicate resolution to data consumer		A		R		I	
7 - Leverage dataset						A/R	

R: Responsible – *Person who does the task*

A: *Accountable – Person who delegates work and provides final review on a task or deliverable before it is deemed complete*

C: *Consulted – Person who provides input on a task or deliverable based on the impact on their work or their domain of expertise*

I: *Informed – Person who needs to be kept in the loop on task or deliverable progress*