

13 Jan, 2026

How to Effectively Classify Assets and Associate Glossary Terms in CDGC

- Krishna V, Principal Product Manager, R&D
- Sohan Dhanuka, Sr Principal Manager, R&D

Today's Speakers



Sohan Dhanuka

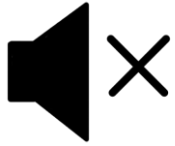
Sr. Principal Product
Manager



Krishna V

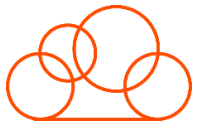
Principal Product
Manager

Housekeeping Tips



- Today's Webinar is scheduled for **1 hour**
- The session will include a webcast and then your questions will be answered live at the end of the presentation
- All dial-in participants will be muted to enable the speakers to present without interruption
- Questions can be submitted to "All Panelists" via the **Q&A option** and we will respond at the end of the presentation
- The webinar is **being recorded** and will be available on our [Success Portal](#) - where you can download the **slide deck** for the presentation. The link to the recording will be emailed as well.
- Please take time to complete the **post-webinar survey** and provide your feedback and suggestions for upcoming topics.

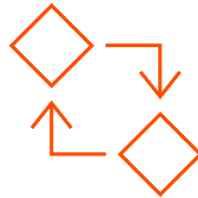
Feature Rich Success Portal



Bootstrap trial and
POC Customers



Enriched Customer
Onboarding
experience



Product Learning
Paths and Weekly
Expert Sessions



Informatica
Concierge



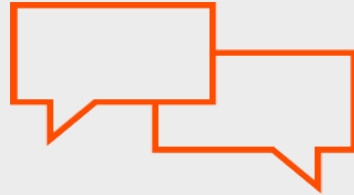
Tailored training and
content
recommendations

More Information



Success Portal

<https://success.informatica.com>



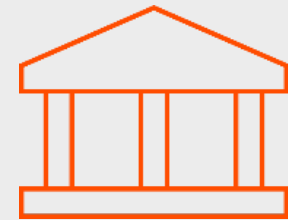
Communities & Support

<https://network.informatica.com>



Documentation

<https://docs.informatica.com>



University

<https://www.informatica.com/in/services-and-training/informatica-university.html>

Safe Harbor

Disclaimer: The information being provided herein is for informational purposes only. The development, release and timing of any Informatica product, service or functionality described herein remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision. Statements made herein are based on information currently available, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products, services or functionality in the future.

Agenda

1 Importance of Classification & Glossary Association in Data Catalog

2 Types of Data Classification & How to configure them in CDGC

3 Intelligent Glossary Association in CDGC

4 Demo

5 Q&A

Importance Classification & Glossary Association in Data Catalog

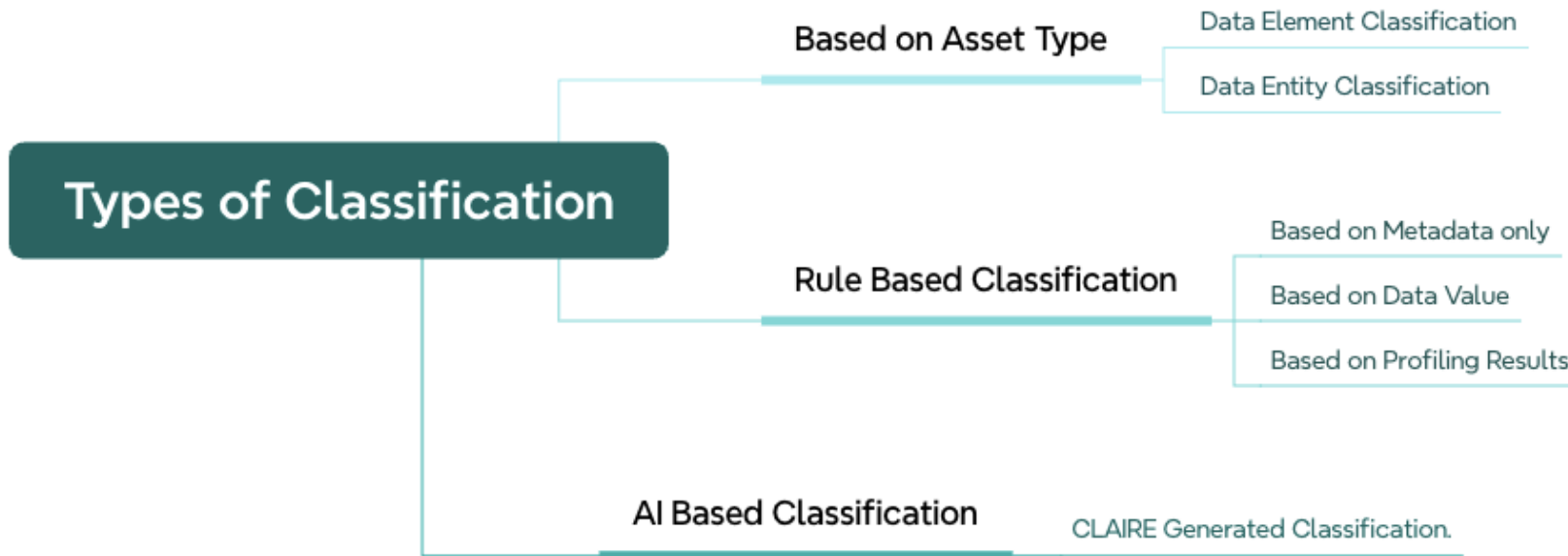
Data Classification

- Enhanced data discovery – Classification & Glossary association adds value to metadata allowing it be organized and easily searchable.
- Data Security – By classifying assets into PII,PHI organization can easily identify assets which are sensitive in nature and take appropriate measures to safeguard it.
- Risk Compliance - Classification allows metadata to be organized as mandated by regulatory needs.

Glossary Association

- Business Context – Glossary provides business context, association to technical assets provides business relevancy & impact.
- Data Literacy & Consistency – Allows non-technical users to interpret data easily and follow uniform terminology across organization and reducing ambiguity.
- Impact Analysis – In enterprise glossary terms are associated with policy, KPI, process etc there which allows better impact analysis in case of any data change.

Types of Classification in CDGC



How to Classify assets?

Rule Based

1. Metadata based – Based on Name of the asset or parent name of the asset.
2. Data based – Based on Value Frequency which matches data pattern or known values
3. Profiling Statistics – Based on Average/Inferred Data Types/Max or Min Values, % of NULL

Attribute ?

Name

Operator

- Appears In
- Appears In Lookup Table
- Appears in Reference 360
- Contains Text
- Ends With
- Is Equal To
- Pattern Like
- Starts With

Attribute ?

Frequent Value

Operator

- Appears In
- Appears In Lookup Table
- Appears in Reference 360
- Contains Text
- Ends With
- Is Equal To
- Pattern Like
- Starts With

Attribute ?

Inferred Data Type

Operator

Is Equal To

Attribute ?

Average

Operator

- Is Equal To
- Is Greater than
- Is Less than
- Is Not Equal To

How to Classify assets?

AI Based

1. Enable “Generated Data Classification” in Catalog Source Configuration in MCC.
2. Promote/Reject – User must either promote or reject the generated classification based on relevance of asset & context
3. Promote Generated classifications to either new classification or to any existing classification

Note: Generated Classification is based on Metadata only.

Choose either CLAIRE generated classifications, data classification rules, or both.

Generated Data Classifications ? Data Classification Rules ?
powered by **CLAIRE™**

Promote Generated Classification

Select a data element classification to associate with the generated classification.

⚠ Before you promote this generated classification, ensure that there are no ongoing bulk import jobs for promoting this generated classification, as this may lead to errors.

Items (358)

Find ⚙

Data Element Classification	Sensitivity	Description	Category
BIC/Swift Code	2 MEDIUM	This data element classification classifies colu...	-
Country ISO Code	3 MEDIUM HIGH	This data element classification classifies colu...	-
deposit code	1 LOW	-	-
Estonia Personal Identification Code	4 PHI	This data element classification classifies colu...	-
International Dialing Code	5 SENSITIVE PHI	This data element classification classifies colu...	-
Italy Fiscal Code	4 PHI	This data element classification classifies colu...	-
Latvia Personal Code	4 PHI	This data element classification classifies colu...	-
Lithuania Personal Code	4 PHI	This data element classification classifies colu...	-
Postal_Code	3 MEDIUM HIGH	Unique code for locations in United States	-
Romania Personal Numeric Code (CNP)	4 PHI	This data element classification classifies colu...	-
USA State Code	1 LOW	This data element classification classifies colu...	-

1 - 25 of 358 Items

< Page 1 of 15 >

Items Per Page: 25

Could not find the data element classification you were looking for? [Click here to create a new data element classification](#)

?

Back Promote Cancel

Additional Capabilities

- Classification Categories – Use classification categories to group element classifications under specific buckets, one classification can be part of multiple categories. Use categories to search assets which are part of classifications.
- Sensitivity Customization – You can customize sensitivity to accommodate additional sensitivity levels.
- Advanced Rule Builder – Use Adv mode to build complex rules which can be based on name/data/profiling results with built in functions and operators. Rule is built on SPARK SQL syntax; you can validate before execution if the rule syntax is confirming SPARK SQL.

Default	Value
<input checked="" type="radio"/>	NONE
<input type="radio"/>	1 Low
<input type="radio"/>	2 Medium
<input type="radio"/>	3 Medium High
<input type="radio"/>	4 PII
<input type="radio"/>	5 Medium Low
<input checked="" type="radio"/>	6 Extreme high
<input type="radio"/>	7 Sensitive PII

Inclusion Rule Validate

Advanced ↶ ↷ ↺

Attributes ↕

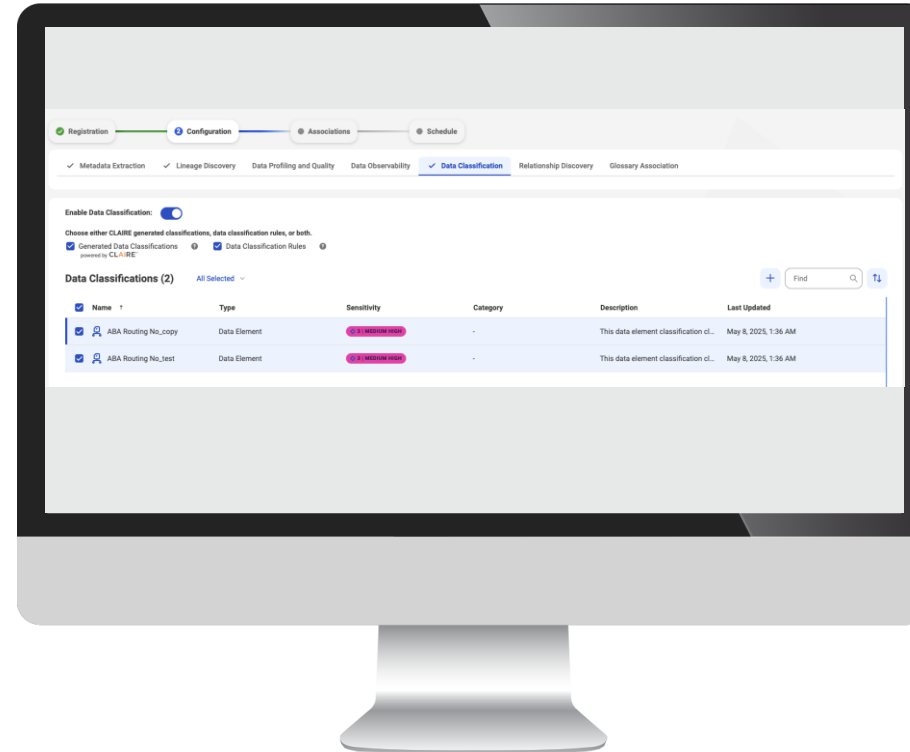
Double-click the attributes or hover on the row to click the arrow and select.

Name	Data Type
Metadata	
COLUMN_COMMENT	String
NAME	String
PARENT_COMMENT	String
PARENT_NAME	String
Statistics	
AVERAGE	Numeric
FREQUENT_VALUES	Array
INFERRED_DATATYPE	String
MAX_VALUE	String
MIN_VALUE	String
NULL_PERCENTAGE	Numeric
NUM_PROFILED_VALUES	Numeric

Select an item from the options on the left to see the description.
Enter your expression in the editor. For example: NAME LIKE "%Credit Card%" AND forall (FREQUENT_VALUES, v -> v RLIKE "[0-9]{12}?-[0-9]{3}?")
To see the list of suggestions in the editor, press Ctrl+Space.

```
AVERAGE == 10.5d AND (size(filter(FREQUENT_VALUES, v -> v IN ('India','US','Japan')))) / size(FREQUENT_VALUES) >= 0.85f
```

DEMO



Glossary Association – Scoring & Threshold

- Glossary Association is based on scoring, CDGC creates a confidence score for every potential match between technical asset and a Glossary term, this score is driven by algorithm.
 - The Score can be used to decide whether to accept or recommend glossary.
 - Users can configure glossary to be accepted if confidence score is between 80-100, else it would be recommended and users must manually accept it.
 - Users can configure glossaries to be recommended based on threshold; Admins can set the threshold between 80-90 to recommend glossaries which can be accepted or rejected by users. If users reject any specific glossaries, those glossaries will not be recommended in next execution phase.
 - Assign Business Name & Description: Users can import business names & description on technical assets from the associated glossaries.

Enable Glossary Association:

Enable auto-acceptance ? Confidence Score Threshold for Auto-Acceptance: * ? 80 100 95 %

Enable Below-threshold Recommendations ? Confidence Score Threshold for Recommendations: * ? 80 95 85 %

How to Improve Glossary Association

Ignore Key Words

Use Ignore key words to define and prefix or suffix which will be ignored by algorithm on technical assets for example to ignore “_c” use suffix and define “_c”, algorithm will exclude “_c” from the technical asset names.

Use Suffix to ignore at the start of technical asset name

Use Prefix to ignore at the end of technical asset name.

Ignore Keywords

Ignore specified keywords.

No Yes

> Hide example

Prefix

Stg_ x Wrk_ x

Suffix

_ID x _tmp x

How to Improve Glossary Association

Glossary Association Scope

Use this option to limit the glossary association to specific hierarchy. Select top level glossary which will be used to limit the glossaries that will be associated with technical asset

Glossary Association Scope

- All Glossary Assets ? Top-level Glossary Assets ?

Removes all existing glossary associations that are not within the scope of the selected glossaries.

Select Top-level Glossary Assets

Select Assets

Top-level Business Glossary Assets (67)

Find



Filter Type: Domain x Add Filter v

Clear All

<input type="checkbox"/> Name ↑	Type	Description	Created By
<input type="checkbox"/> Account Customer Segme...	Domain	Account Customer Segme...	Sandeep Suresh (Inact... i)
<input type="checkbox"/> Account Customer Segme...	Domain	Account Customer Segme...	Sandeep Suresh (Inact... i)
<input type="checkbox"/> Account Customer Segme...	Domain	Account Customer Segme...	Sandeep Suresh (Inact... i)
<input type="checkbox"/> Adventure Works	Domain	The Adventure Works data...	Sohan Dhanuka
<input type="checkbox"/> Balance OPS	Domain	Balance operation	Sandeep Suresh (Inact... i)
<input type="checkbox"/> Balance OPS	Domain	Balance operation	Sandeep Suresh (Inact... i)
<input type="checkbox"/> CDomain	Domain	-	Sreedhar Panduri
<input type="checkbox"/> City	Domain	City testing EDC Business ...	Sandeep Suresh (Inact... i)
<input type="checkbox"/> Customer Churn	Domain	Customer attrition, also kn...	Sandeep Suresh (Inact... i)

1 - 25 of 67 Items

Page 1 of 3

Items Per Page: 25



OK

Cancel

How to Improve Glossary Association

Abbreviation & Synonym

Create & use lookup tables to add any synonym or abbreviations to improve glossary associations, for example INV can be abbreviated to “Invoice” which be improve scoring for assets which contain “INV” in technical asset names

Use Abbreviation and Synonym Definitions

Yes No

lkp_invoice

Select

General Information

Name: * ?

lkp_invoice

Identifier: ?

lkp_invoice

Description:

Enter a description for the lookup table

Code Page for Delimited Files: ?

UTF-8

Glossary Association Synonyms: ?



Lookup Table File: * ?

Select a CSV file

Browse

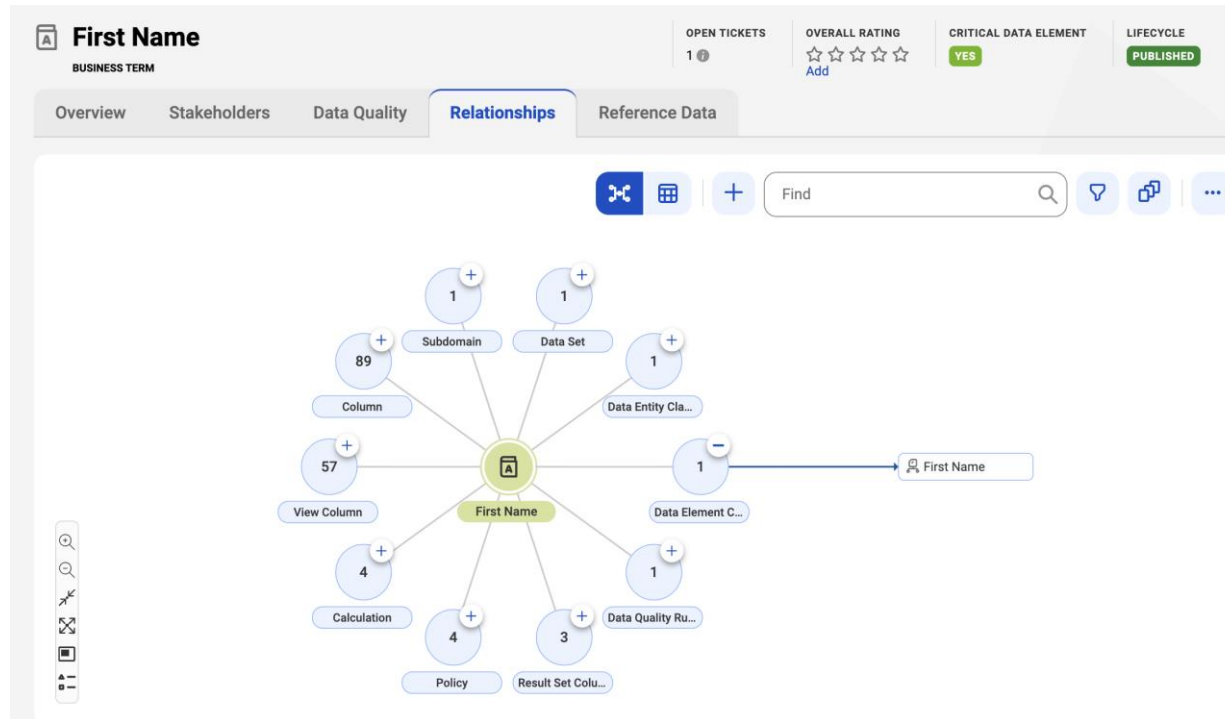
Hide Lookup Table File Restrictions

- The file size must not exceed 5 MB.
- The file must contain Abbreviation and Synonym column headers. Use a comma as the delimiter
- The first line in the file must be a header based on which column names are identified.
- Column names can contain letters, digits, and underscores. They cannot contain special characters. ASCII punctuation symbols such as @ # & ? and ~ are not allowed.
- The column size must not exceed 150 characters.
- The row length must not exceed 750 characters.

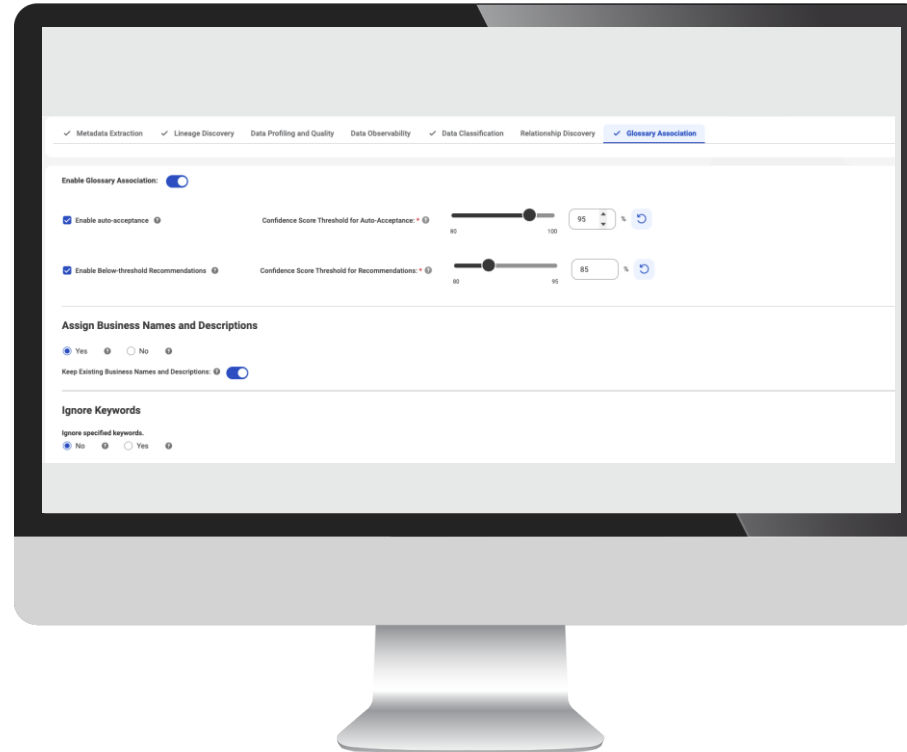
How to Improve Glossary Association

Create Relationship between Glossary & Classification

Based on classification association with assets Glossary is automatically assigned to technical assets



DEMO



KB Articles for Additional Details

Data Classification with conformance & weighted conformance score

https://knowledge.informatica.com/s/article/CDGC-Data-Classification-Rules-Conformance-vs-Weighted-Conformance?language=en_US

Test Classification Rules on SPARK Clusters.

https://knowledge.informatica.com/s/article/How-To-Test-CDGC-Classification-Rules?language=en_US&type=external