

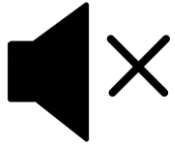
07 Nov, 2024

# What's New in Informatica Cloud Data Governance and Catalog (CDGC), Cloud Data Marketplace (CDMP) and Cloud Data Quality (CDQ) in October and November 2024 Release?

- Amrish Thakkar, Senior Director, Product Management, R&D
- Kate Amory, Director, Product Management, R&D
- Ricardo Sobral, Senior Director, Product Management, R&D

Where data  
& AI come to 

# 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.

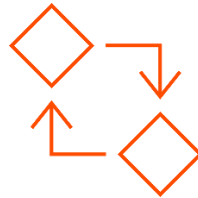
# Feature Rich Success Portal



**Bootstrap trial and  
POC Customers**



**Enriched Customer  
Onboarding  
experience**



**Product  
Learning Paths  
and Weekly  
Expert Sessions**



**Informatica  
Concierge**



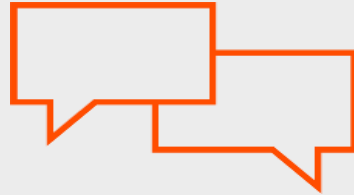
**Tailored training  
and content  
recommendations**

# More Information



## Success Portal

<https://success.informatica.com>



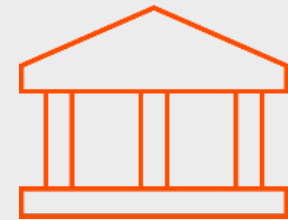
## Communities & Support

<https://network.informatica.com>



## Documentation

<https://docs.informatica.com>



## University

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

# Safe Harbor

The information being provided today is for informational purposes only. The development, release, and timing of any Informatica product or functionality described today remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision.

Statements made today are based on currently available information, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products or functionality in the future.

# Cloud Data Governance and Catalog

# Features in CDGC 2024.11.M

- Metadata Access Control
- Dashboard/Saved Search sharing
- Dashboard Improvements
- Business Process Visualization
- Search enhancements
- Export and Import changes
- Public APIs - support for CSV
- Claire powered Inferred Lineage
- Usability updates
- QuickLook
  - Support for French/Spanish/Brazilian Portuguese
- Discovery & Profiling updates
- Scanner Additions and enhancements

# Metadata Access Control



# Metadata Access Control



## Key Highlights

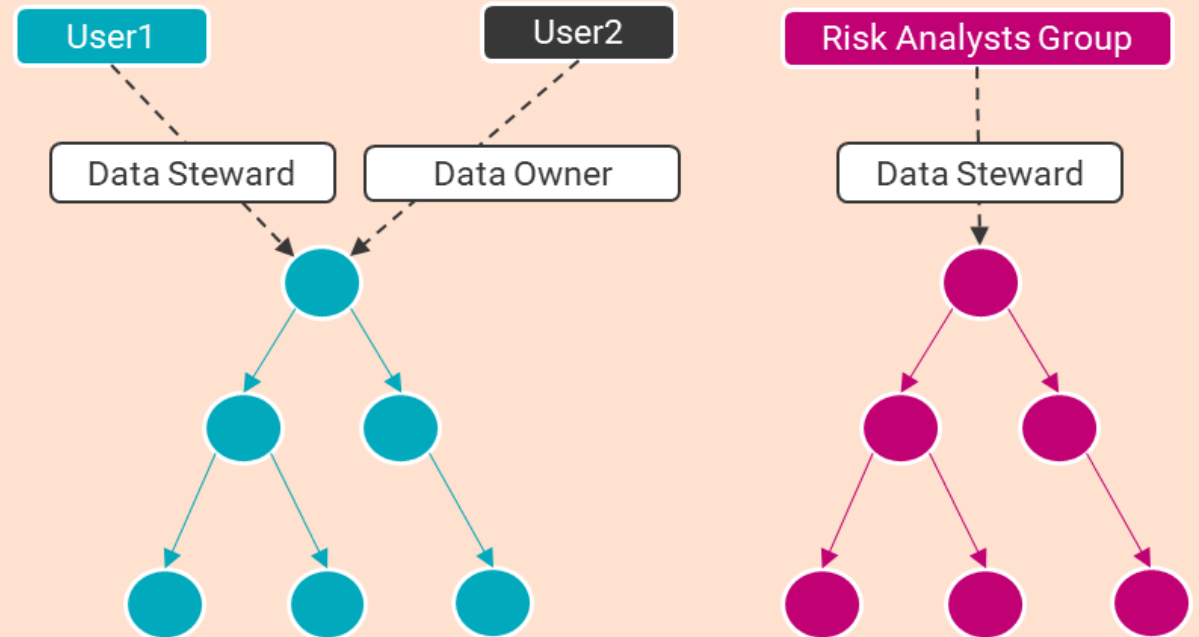
- **Easy and efficient metadata access control** extending IDMC user roles definition
- **Attribute value driven metadata access control** for all asset at any level of the hierarchies
- **Granular metadata access control** for any asset of the metadata platform via asset specific roles (stakeholders) definition



## Benefits

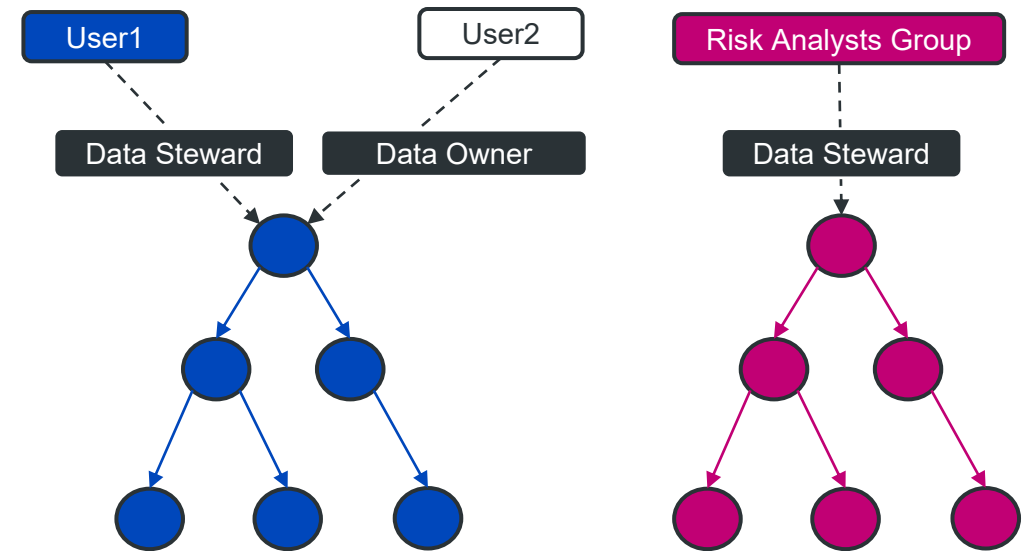
- **Provide full control** to customers to control access to dataset information at an enterprise level
- **Allow customers to catalog the entire data landscape** securely and manage asset ownership at scale
- **Allow customers to open consuming application (CDGC, CDMP, CLAIRE GPT)** to their entire organization

## Metadata Access Control



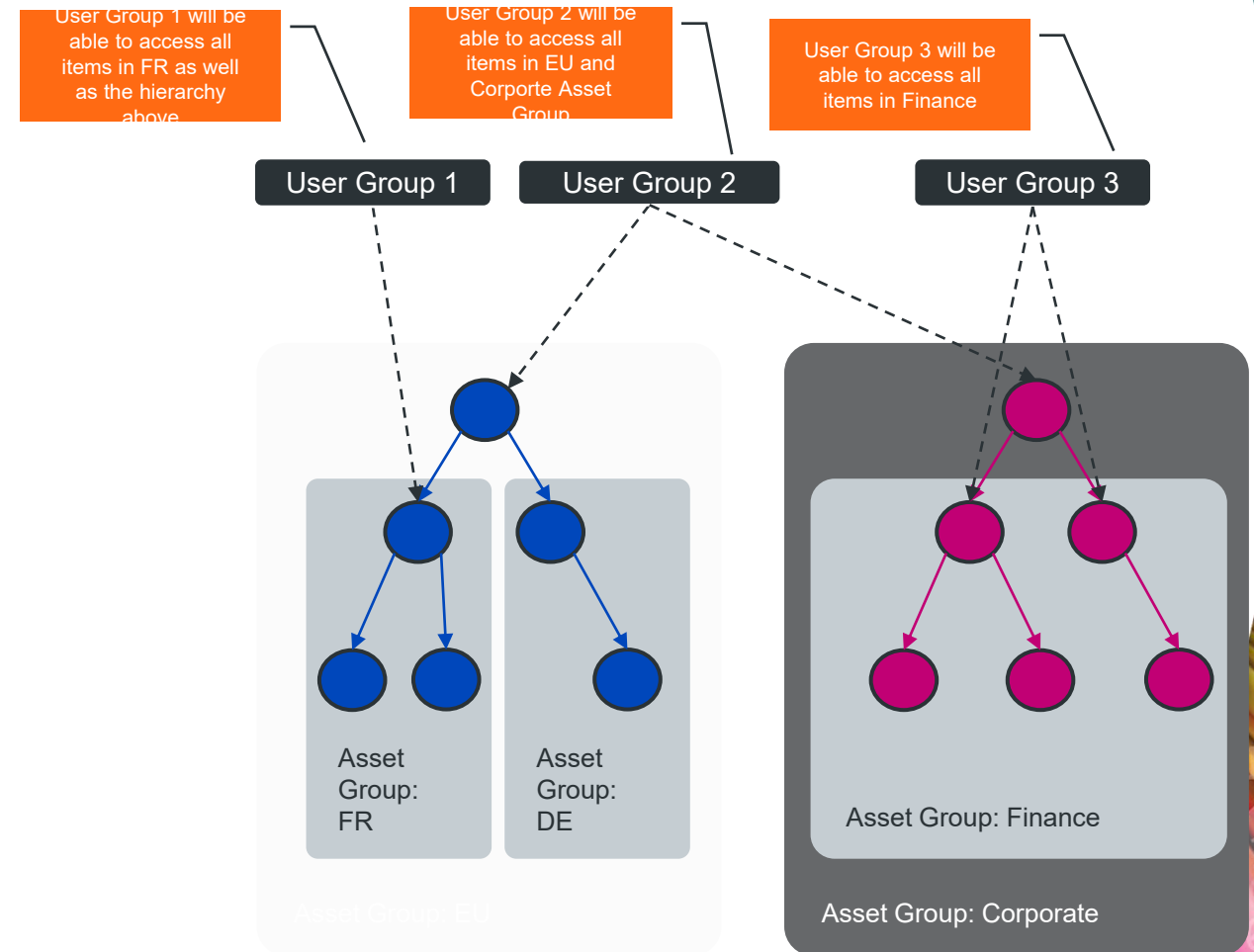
# Enhanced control of permissions based on roles

- **User roles** grants access to asset types based on their roles.
  - Permissions can be granted at the level of abstract types (Business assets, Marketplace Assets and Technical assets).
  - Permissions can be granted at individual type level
  - As model evolves, new asset types can be introduced into the system dynamically.
  - Technical assets: Different access controls can be defined for the same class type based on catalog source type.
- **Stakeholder roles** are used to control who gets to edit or make changes to an asset.
  - Presence of a stakeholder on an asset prevents non stakeholder users to perform permission granted through user roles on the asset.
  - In case of Tech Asset, Stakeholder assignment are inherited from parent asset unless overridden.
- Roles can also drive access to certain information on the Asset:
  - **Profiling, data, code, unpublished changes**



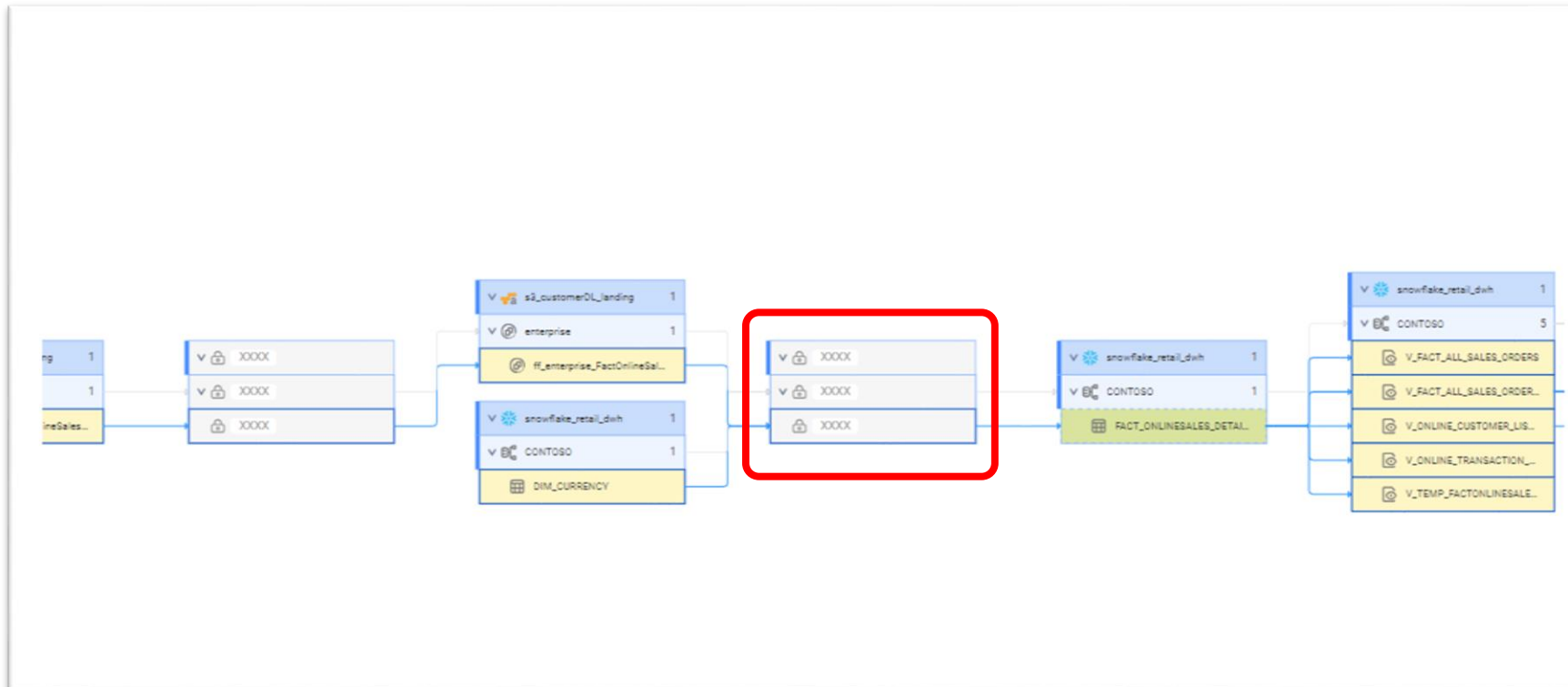
# Control visibility of assets based on attribute

- Access to assets can be restricted based on attribute value
  - In November, Access control based on **lifecycle, catalog source type, reference flag and Asset Group**
  - Lifecycle: Visibility based on Lifecycle values
    - For published asset, access to ongoing changes can be restricted as well
  - Catalog source type and reference: to reduce the scope permissions on technical assets
- Asset Group: Parameters on an Asset that drives visibility of the asset for the user
- A user should have a permission on asset type to gain that permission on an asset through attribute-based access control.



# Access Control on Lineage & Relationship View

- Assets on the lineage/relationship view that the user does not have access to will be masked out.
- This is to ensure users can see the complete lineage with certain items being masked out based on user's access to those assets



# Metadata Access Control – Use Cases

## Hierarchy based Access Control

- Restrict access on hierarchies across business and technical assets to align with company's organization and responsibility model
- Restriction can be added at any level in the hierarchy i.e. does not always has to be at the Root
- For example, restricting access to certain Glossaries or Catalog Sources for a User

## Lifecycle based Access

- Allows to hide drafts and unpublished changes for non editor users
- Reduces noise for the end users as they see published/approved assets only

## Persona driven Access

- Allows access to the assets based on the user persona
- For example, Hide/Restrict access to Technical Assets for Business Users.
- This improves user experience by only showing relevant and contextual information to the user

# Dashboard and Saved Search Sharing

# Dashboard Sharing

- **Peer-to-Peer Dashboard Sharing:**  
Any user can share their dashboards with others
- **Flexible Sharing Options:**
  - Share dashboards with **users, groups, or roles.**
  - Assign permissions to control access, allowing others to either **view** or **edit** the shared dashboards
  - Share multiple dashboards
  - Share Saved Searches used in dashboard

### Manage Sharing

Private  Share with specific users, user groups and roles  Share with all users

**i** Select one or more users, user groups, or roles to share the dashboard with, and assign them permissions to view or edit this dashboard.  
You can also set this dashboard as the default one. If you have earlier provided another default dashboard, 'Admin Dashboard' will replace that dashboard. If users have set their own dashboard as default, they will remain as default.

Select:

**Users** **User Groups** **Roles**

Share searches used in the dashboard widgets **i**

Shared with (8) **i** Double-click rows to edit

Name	Email ID	Permissions	<input type="checkbox"/> Set as Default Dashboard <b>i</b>
Governance User	████████@informatica.com	View only	<input type="checkbox"/>
Admin	-	Allow Editing	<input type="checkbox"/>
API Policy Manager	-	<input type="button" value="Allow Editing"/>	<input type="checkbox"/>
Business Manager	-	<input type="button" value="View Only"/>	<input checked="" type="checkbox"/>
CLAIRE GPT User	-	<input type="button" value="Allow Editing"/>	<input checked="" type="checkbox"/>
Customer 360 Analyst	-	Allow Editing	<input checked="" type="checkbox"/>
Customer 360 Data Steward	-	Allow Editing	<input checked="" type="checkbox"/>
Customer 360 Manager	-	Allow Editing	<input checked="" type="checkbox"/>

# Dashboard Sharing

- **Centralized Dashboard Management**

- **All Dashboards:** View and manage all your dashboards in one place
- **Shared by Me:** Keep track of dashboards you have shared with others
- **Shared with Me:** Access dashboards that have been shared with you
- **Transfer ownership:** Users with Manage Access Control privilege can transfer ownership of inactive or deleted users given the dashboards were shared with these users

The screenshot shows the 'Dashboard Sharing Manager' interface with the 'All Dashboards' tab selected. It displays a table with 6 dashboards. The table has columns for Name, Shared With, Shared By, Permissions, Set as Default For, Owner, and Updated On.

Name	Shared With	Shared By	Permissions	Set as Default For	Owner	Updated On
Admin Dashboard	1 Users	You	View Only (1)	-	Achal Shah	Oct 27, 2024, 12:...
Clone My Dashboard	-	Sreedhar Panduri TEAM	Edit	-	Sreedhar Panduri TEAM	Oct 27, 2024, 1:2...
Governance Dashboard	1 Roles	You	View Only (1)	-	Achal Shah	Oct 27, 2024, 11:...
My Dash	1 Users	You	View Only (1)	-	Achal Shah	Oct 27, 2024, 12:...
New 1	-	Sreedhar Panduri TEAM	Edit	-	Sreedhar Panduri TEAM	Oct 27, 2024, 12:...
Policy	1 Roles and 2 Users	You	Allow Editing (3)	-	Achal Shah	Oct 27, 2024, 3:0...

The screenshot shows the 'Dashboard Sharing Manager' interface with the 'My Dashboards' tab selected. A context menu is open over the 'Permissions' column of the 'Admin Dashboard' row. The table has columns for Name, Shared With, Shared By, Permissions, Set as Default For, Owner, and Updated On.

Name	Shared With	Shared By	Permissions	Set as Default For	Owner	Updated On
Admin Dashboard	1 Users	You	View Only (1)	-	Achal Shah	Oct 27, 2024, 12:...
Governance Dashboard	1 Roles	You	Edit Dashboard (1)	-	Achal Shah	Oct 28, 2024, 6:3...
My Dash	1 Users	You	Delete Dashboard (1)	-	Achal Shah	Oct 27, 2024, 12:...
Policy	1 Roles and 2 Users	You	Share Dashboard (3)	-	Achal Shah	Oct 27, 2024, 3:0...



# Saved Search Sharing

- Any user creating a saved search can choose to share with specific users, user groups, or roles.
- Saved Searches can be shared with viewer or editor permissions.
- Users can view both shared and private saved searches on the Saved Searches page.
- When sharing a dashboard, the owner can choose to share saved searches used in widgets via a “Share Saved Searches” option, which is enabled by default.
- Searches shared with the dashboard are shared with Viewer-only permissions.

The image shows two screenshots from an application. The top screenshot is a dialog titled "Share this Saved Search". It has three radio buttons for sharing options: "Private", "Share with specific users, user groups and roles" (which is selected), and "Share with all users". Below this are tabs for "Users", "User Groups", and "Roles". A search input field contains "Type a Name or Email...". To the right is a "Permission:" dropdown menu set to "View Only" and an "Add" button. Below the input is a table with the instruction "Double-click rows to edit." The table has columns for "Name", "Email", and "Permission".

Name ↓	Email	Permission
Karl Thompson	kathompson@informatica.com	Allow Editing
Person Name	personname@emaildomain.com	View Only
Person Name	personname@emaildomain.com	View Only
Group Name (12)	-	View Only

The bottom screenshot is the "Saved Searches" page. It has tabs for "All Saved Searches (14)", "My Saved Searches (6)", and "Shared with Me (8)". The "Shared with me (8)" tab is active. It shows a list of saved searches. The first one is "Selected Saved Search Name" with a description and filters: "Asset Type: Business Term", "Stakeholder: John Doe", "Timeframe: Last 90 Days". It was updated on Aug 20, 2024. The second one is "Critical Data Element Profiled 5 Stars" with filters: "Asset Type: Data Element", "Status: Critical", "Rating: 5 Stars". It was created on Aug 16, 2024.

# Dashboard Improvements

# Dashboard Improvements

## Custom Attributes in List View of Search

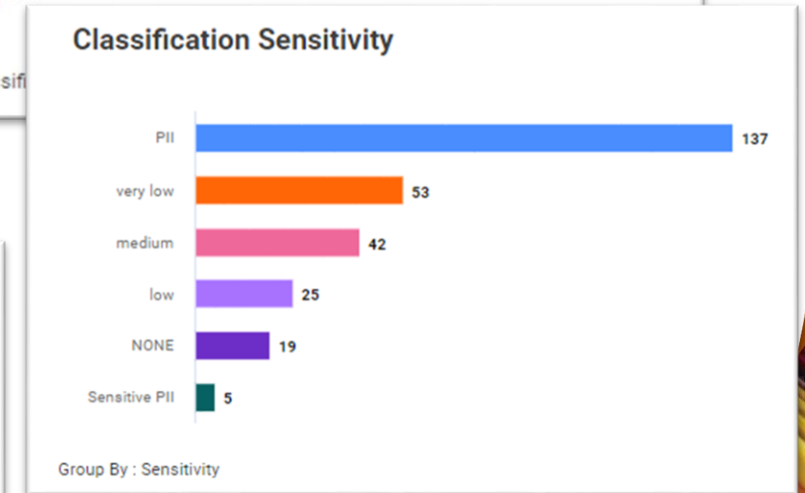
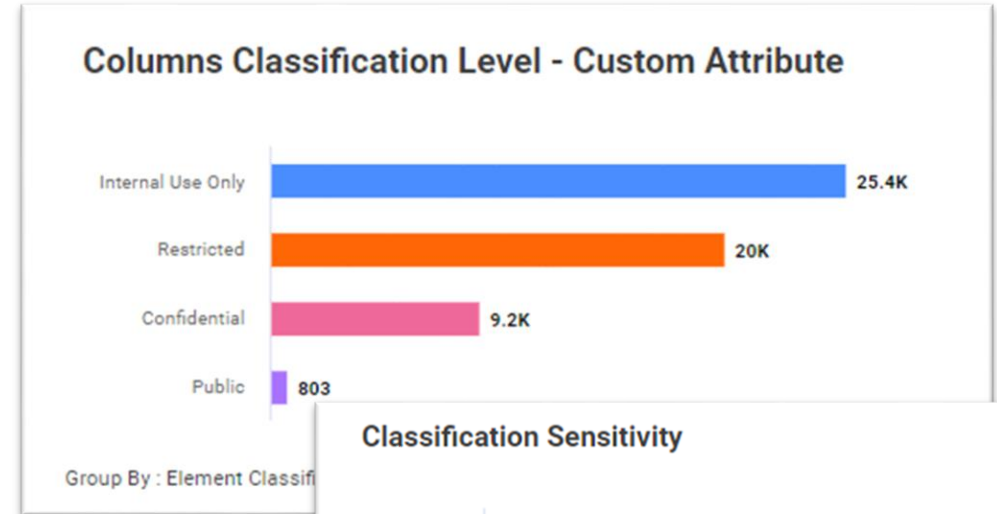
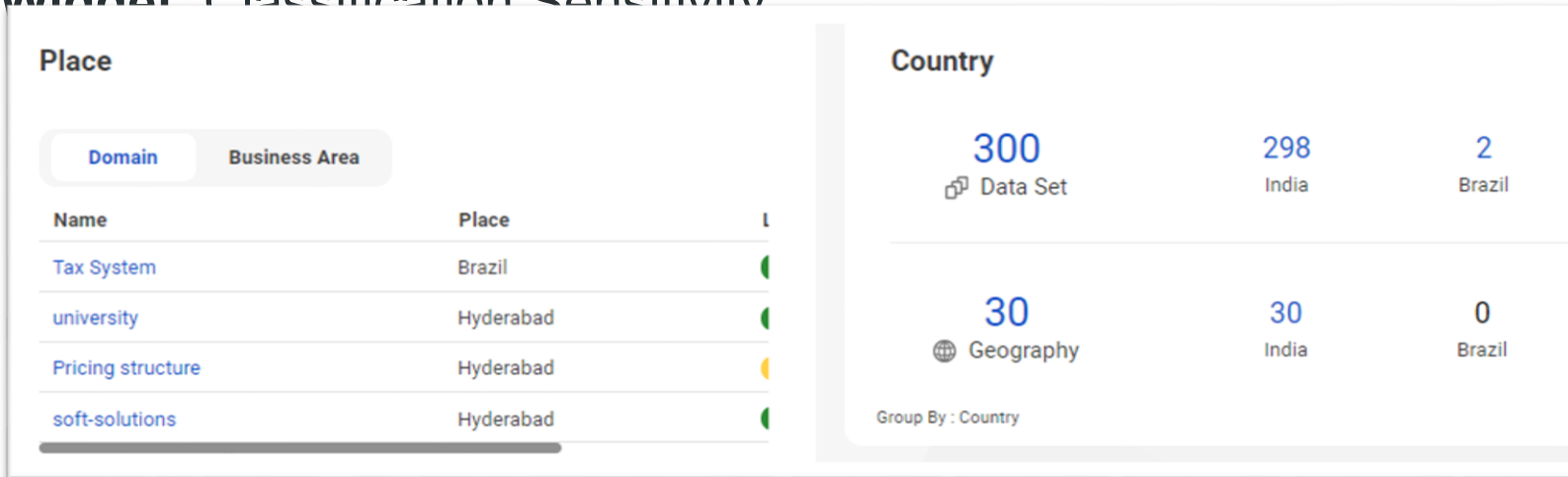
**Widgets:** Users can now select and display custom attributes within the list view of dashboard widgets.

## Custom Attributes in Group By of Search

**Widgets:** Users can now group assets by custom attributes.

## Additional Attributes in Group By of Search

**Widget:** Classification Sensitivity

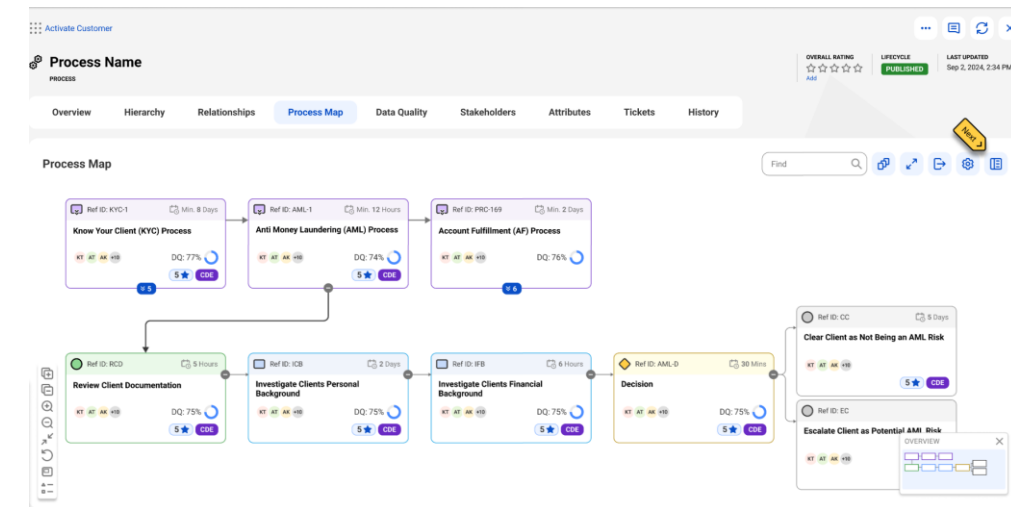
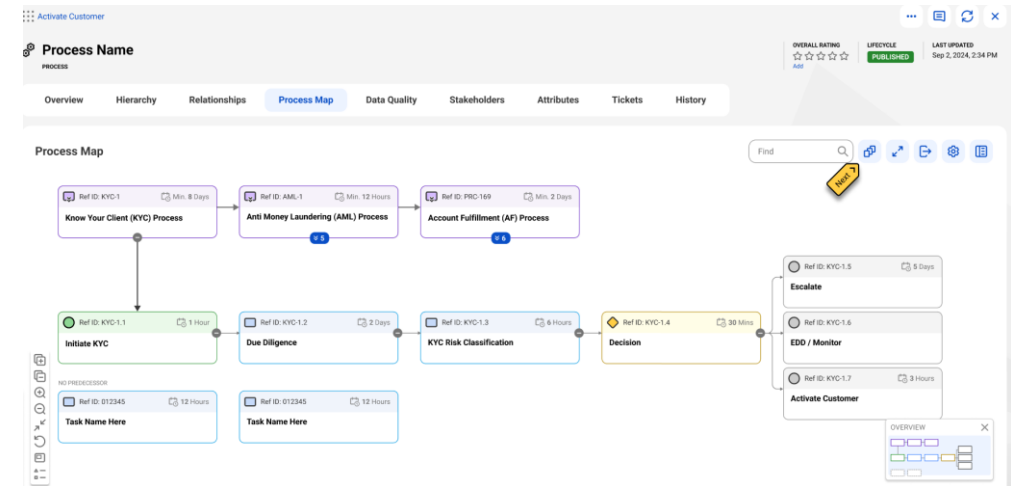


# Business Process Visualization

# Process Flow Maps

Visualize the order in which a process is executed

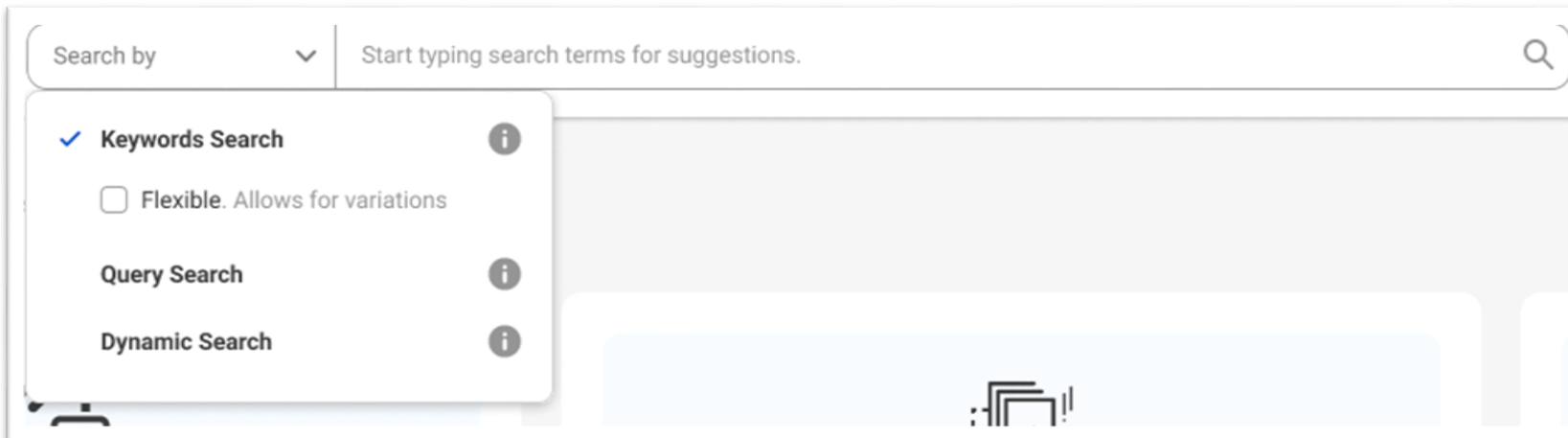
- New Tab in Process Asset: Process Map
- Powerful controls for exploring larger Maps
  - Expand / Collapse sections
  - More space for Process Step names
  - Overlays
  - Zoom In/Out
- No Model change required – uses the information you have already added



# Search Changes

# Search selection

- User will now be able to select the type of Search they want to execute
  - Selection will be between Keyword/Text search, Query Search or Dynamic search (where system picks between keyword and query)
  - For Keyword search, user will be able to go between Exact or Flexible/Fuzzy search
- User selection will be saved automatically so next time they start with that type
- Saved Searches can still be executed irrespective of the Search type selection



# Other Search changes

- Exact count shown instead of showing “10K+”
  - This will only apply to non relationships based searches
- Begins with (abc\*) and Ends with (\*xyz) search will strictly return items beginning/ending with that sub string
  - This will only apply to Name, Business Name and Alias Name fields



# Export and Import improvements

# Export technical assets on the Catalog Source page

Users have two options to export assets directly from the catalog source page -

- **Export Catalog Source:** Exports all assets in a catalog source, with "Include Hierarchy" selected by default.
- **Export Specific Assets:** Allows export of specific assets by selection of asset types and specific assets.

## Additional Filter Options

- **Assets with Pending Recommendations:** Users can filter and export only those assets requiring curation on pending glossary recommendations
- **Stakeholder:** Users can filter assets by assigned stakeholders or assets without stakeholders
- After filters are applied, users can choose to include or exclude hierarchy, and other asset details
- If asset selection or filtering results in zero asset count, the Export button is disabled.

The screenshot shows the 'Select Assets' interface. At the top, there are two tabs: '1 Select Assets' and '2 Export'. Below the tabs, there are radio buttons for 'Export: Asset Types' and 'Specific Assets', with 'Specific Assets' selected. A search bar contains the text 'Start typing search terms to see suggestions.' Below the search bar, it says 'Items (8) 4 Selected'. A table lists items with columns for Name, Type, Reference ID, and Hierarchy. The first seven rows are 'Table' type assets with Reference IDs BT-123, BT-345, BT-212, BT-321, BT-421, BT-561, and BT-654. The eighth row is a 'Database' type asset. The first four rows have checkboxes checked, indicating they are selected. To the right of the table, there is a sidebar with a message: 'Select an asset from the left pane to view asset details.'

Name	Type	Reference ID	Hierarchy
Asset Name 123456...	Table	BT-123	Top Level
Asset Name 123456...	Table	BT-345	Top Level
Asset Name 123456...	Table	BT-212	Top Level
Asset Name 123456...	Table	BT-321	Top Level
Asset Name 123456...	Table	BT-421	Top Level
Asset Name 123456...	Table	BT-561	Top Level
Asset Name 123456...	Table	BT-654	Top Level
Asset Name 123456...	Database	-	Teradata / ... / UTF8_ALL_DATA

The screenshot shows the 'Export Catalog Source' dialog box. It has a title 'Export Catalog Source'. Below the title, there is a section for 'Additional Filters' with the text 'Applying these filters may change the number of assets being exported.' There are two checkboxes: 'Assets with Pending Recommendations only' (unchecked) and 'Stakeholders: Select' (selected). Below the 'Stakeholders' section, there is a dropdown menu with 'Select' in it. At the bottom, it says 'Exporting: 10 Asset Types, 1234 Assets' and there is a checkbox for 'Include Asset Details' (unchecked).

# Bulk Import Enhancements

## Multi-Port Rules supported for Data Quality Rule Occurrence Bulk Import Template

- Multi-port rule occurrences created via DQRO Bulk Import template now support port mapping.
- Port Name is appended as the final element in the Secondary Data Element path in the format: CSNAME://DBNAME/SCHEMA/TABLE/COLNAME:PORTNAME

## Added Primary and Secondary Glossary columns to the Data Quality Rule Template bulk import template

- Users can add glossaries (Business Term or Metric) that are in Published state to Primary and Secondary glossary columns in the DQRT bulk import template.

## Support creation of Total Runs for scorecard ROs through DQRO bulk import template

- When users create or update Rule Occurrences with scores, a new child object, 'Total Runs,' will be created with the attributes: Scanned Time, Total Rows, Failed Rows, and Latest Score.
- Updating non-score attributes (e.g., description) via Bulk Import will not create an additional Profile Run Object

# Other Enhancements

# Public API – CSV Support

- **Bulk Import** and **Bulk Export** Public API now supports **CSV** files

`{{api.url}}/data360/search/export`

`{{api.url}}/data360/content/import`

- Additional Optional Parameter for **export**

- `filetype` : CSV / EXCEL
  - Default value is EXCEL

- Export will be in zip file containing multiple CSV files, one file for each asset type.
- No additional parameters for import, (File format is automatically detected).

The screenshot displays the 'Export\_23.csv\_export' job status page. The job is marked as '100% Completed'. The 'Job Summary' table lists the following data:

File Name	Sheet	Asset Type	Exported Assets
Catalog Source.csv		Catalog Source	6
Technical Data element.csv		Technical Data element	9628
Technical Data Set.csv		Technical Data Set	112
Data Set.csv		Data Set	3
Data Element Classification.csv		Data Element Classification	1
Business Area.csv		Business Area	1
Business Term.csv		Business Term	60
Parameter.csv		Parameter	1
Custom.csv		Custom	7

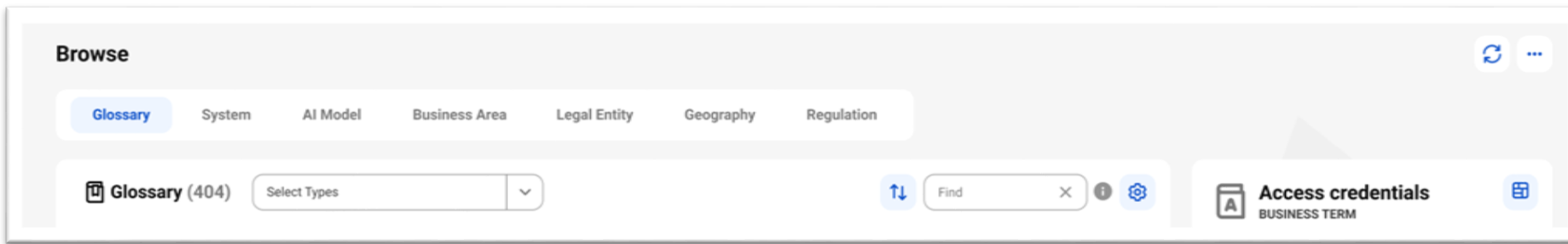
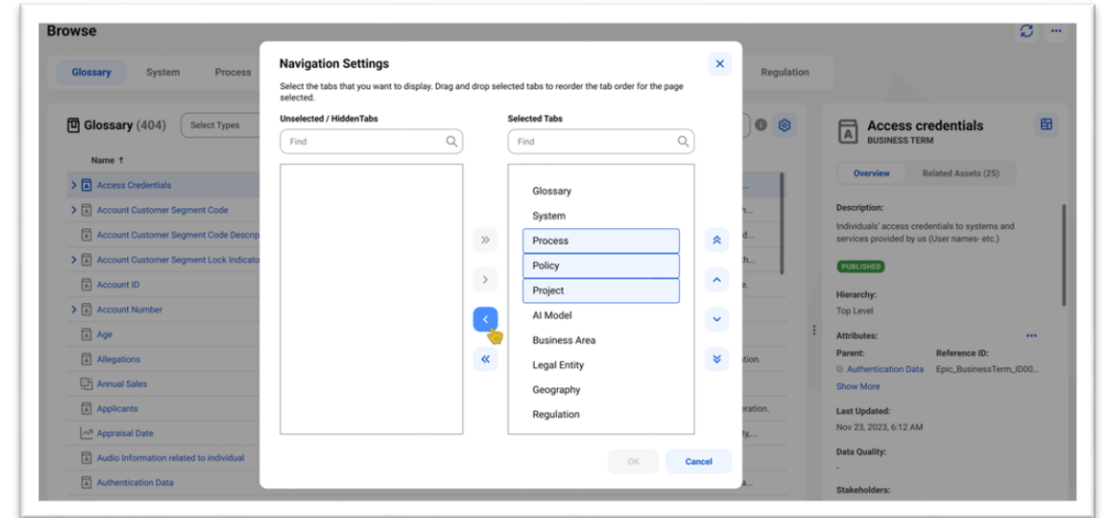
The 'Job Details' section on the right provides the following information:

- Job ID:** 8c15c4cb-5c8d-4a51-814a-9bec...
- Job Type:** Export
- Start Time:** Oct 28, 2024, 12:16:43 AM
- End Time:** Oct 28, 2024, 12:30:1
- Trace ID:** ccgf-searchv2-5-3-0-60867-7fb7...
- Execution Duration:** 00:13:36

# Browse Menu Customization

## Asset Type Tab Customization on Browse Page

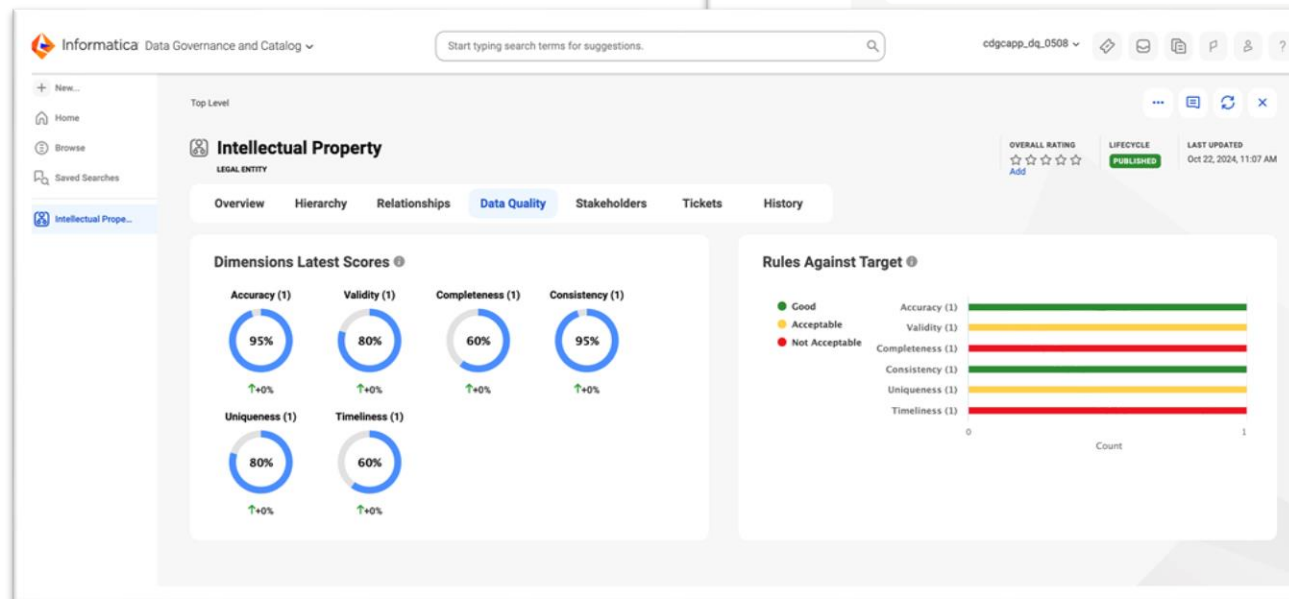
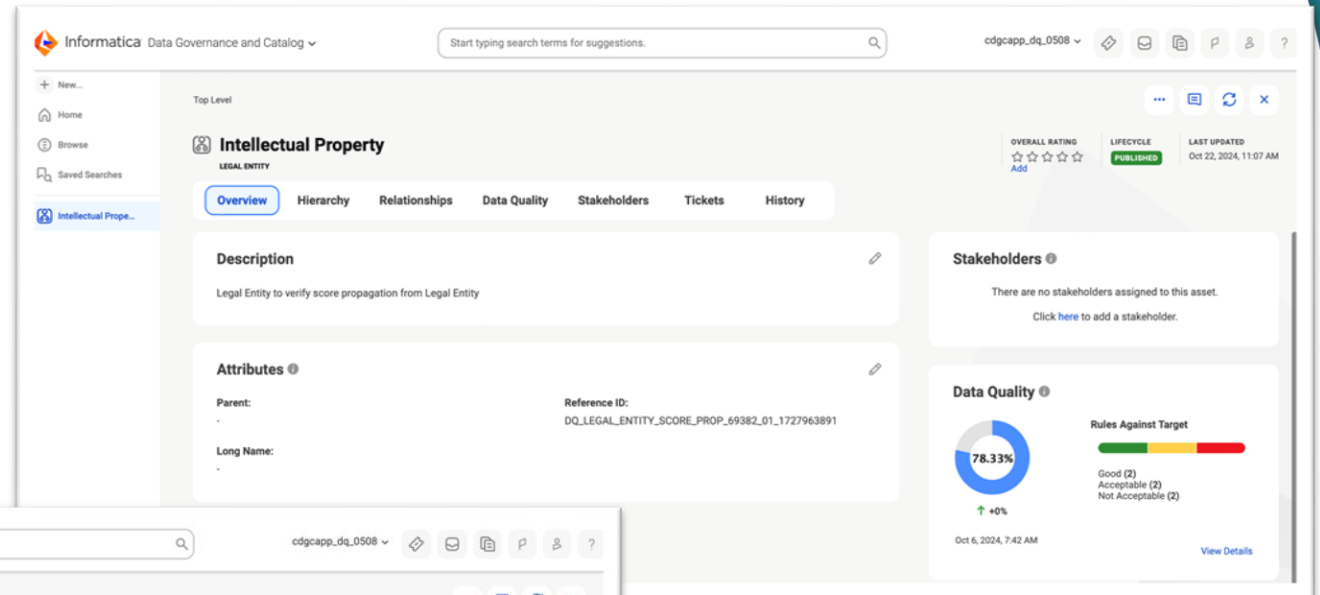
- Users can hide asset type tabs they are not interested in directly on the Browse Page, creating a personalized view.
- Users can select or unselect tabs based on their preferences and save these choices.
- Preferences are retained, allowing users to focus on relevant asset types each time they browse.
- At least one asset type tab must be selected before saving preferences in the settings dialog.



# DQ Changes

## DQ Score for Legal Entity and Business Area assets

- Data Quality score aggregation is now available for Legal Entity and Business Area asset
  - It aggregates through related System and Datasets



# CLAIRE powered Catalog Source Linking (Inferred Lineage)

- CLAIRE powered linking: Uses **trained AI** models to automatically detect and build lineage between user provided source and target
  - Dataset level
  - Data Element level
- Lineage is **recommended** or **auto-accepted** based on user defined option / threshold
- Recommended lineages are visible to stakeholders
- Accepted dataflows are visible as lineage in CDGC application
- Allows user curation (Accept/Reject) of generated lineages

\* Supported for Relational Sources. Will be extended to other sources in future release.



CLAIRE recommends and builds lineage for dataflows happening through manual process, custom inhouse solutions, non-supported ETL, programming languages or any data replication tools.

The screenshot shows a web-based configuration interface for the 'Linking Method'. At the top, there are three tabs: '1 Registration', '2 Configuration', and '3 Linking Method', with the third tab being active. Below the tabs, the 'Linking Method' section is displayed. It includes a 'Method' section with two radio buttons: 'Rule Based Linking' (unselected) and 'Automated Linking powered by CLAIRE™' (selected). The 'Refresh Frequency' section has a checked checkbox for 'Whenever an update is detected for source and/or target catalog source (Recommended)'. The 'Enable Auto Acceptance' section has a toggle switch that is turned on. The 'Confidence score threshold for auto acceptance' section features a horizontal slider ranging from 80 to 100, with a blue marker set at 95. A text input field next to the slider shows '95 %' and a refresh icon is located to the right of the input field.



# CLAIRE powered Catalog Source Linking (Inferred Lineage)

**Configurations (6)**

Find Configuration

**New Linking**

Source: EDW\_Retail\_Prod.HAWK1  
Target: Snowflake\_Retail.Contoso  
Method: Automated Linking powered by CLAIRE™

**Configuration 1**

Source: Snowflake\_Detail.Contoso2  
Target: EDW\_Retail\_Prod.HAWK1  
Method: Rule Based Linking

**New Linking**

TOTAL LINKS: 2534    ACCEPTED: 2530    RECOMMENDED: 4    REJECTED: 0

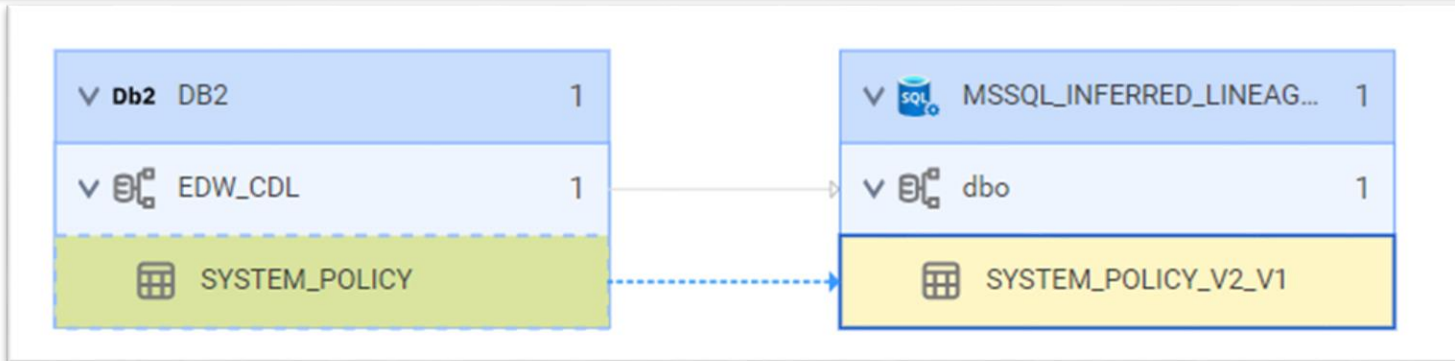
1/2534 Selected    **Accept**    **Reject**    ⓘ

Source	Target	Confidence Score	Status
[-] Root	[+] Root	98%	Auto-Accepted
<input checked="" type="checkbox"/> Leaf	[+] Leaf	89%	Recommended <span>⋮</span>
<input type="checkbox"/> Leaf	[+] Leaf	94%	Recommended
<input type="checkbox"/> Leaf	[+] Leaf	99%	Auto-Accepted
<input type="checkbox"/> Leaf	[+] Leaf	88%	Recommended
<input type="checkbox"/> Leaf	[+] Leaf	96%	Auto-Accepted

Accept

Reject

View Lineage



# Discovery and Classification PAM

- Classification and Glossary Association has been enabled for the following sources:
  - Informatica Power Center
  - Qlik View
  - Qlik Sense

# Scanner updates

# Scanners Updates 2024.11

## New Scanners

- Oracle Data Integrator
- Database SQL Scripts Scanner:
  - Google BigQuery
  - Hive QL
- SAP Analytics Cloud (Preview)
- SAP Datasphere (Preview)
- HDFS (on Dataproc)
- SAS Base (IPS Solution)

# Scanners Updates 2024.11

## Scanner Enhancements

- **Incremental Scanning:** AWS S3, Snowflake
- **Snowflake:** Snowpark Python Stored Procedures
- **Databricks**
  - Support for complex datatype
  - Unity Catalog lineage from system tables with Notebooks references
- **Informatica IICS**
  - CDI ELT support
  - Support for additional CDI components: Data Masking transformation, enteredQuery\_SnowflakeCloudDataWarehouseV2 transformation, Microsoft Fabric Lakehouse\_table and FabricDW\_table
- **Lakes:** various file structure detection improvements

# Scanners Updates 2024.11

## Scanner Enhancements

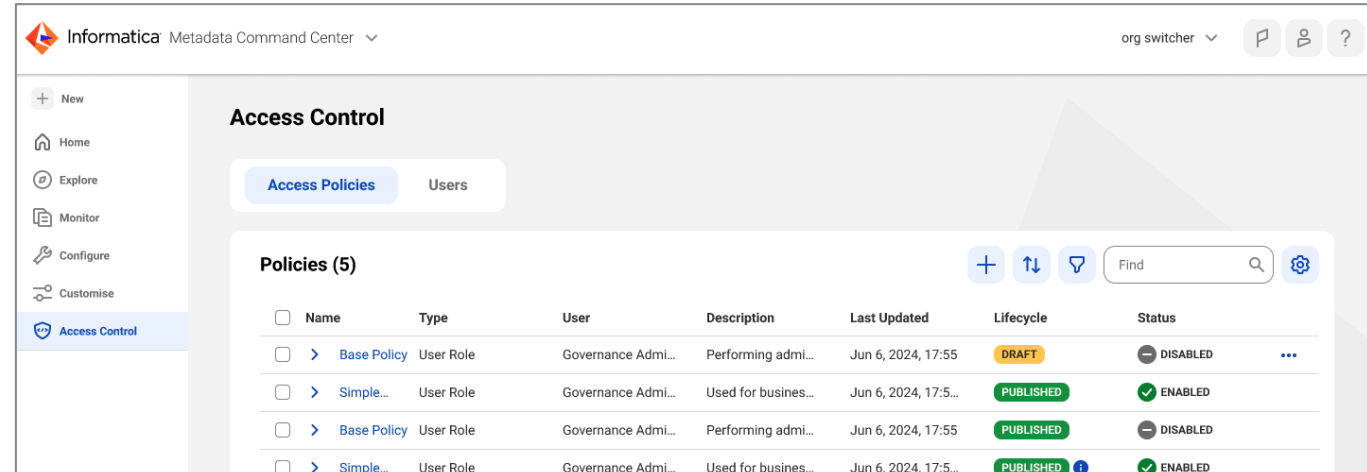
- **Tableau:** external link to the source report
- **Microsoft Azure Data Factory** support for additional activities:
  - System variables
  - Filter
  - Fail
- **Microsoft PowerBI**
  - Support for Fabric sources
  - Dataset structure amendments (Dataset->Tables->Columns)
- **Microsoft Excel:** enhanced support for large files

# Marketplace - November Release

Where data & AI come to **LIFE**

# Summary of Features in Marketplace - November

- Metadata Access Control
- Custom Roles
- More Asset Types from CDGC
- Usability Enhancements
- API Improvements



Informatica Metadata Command Center

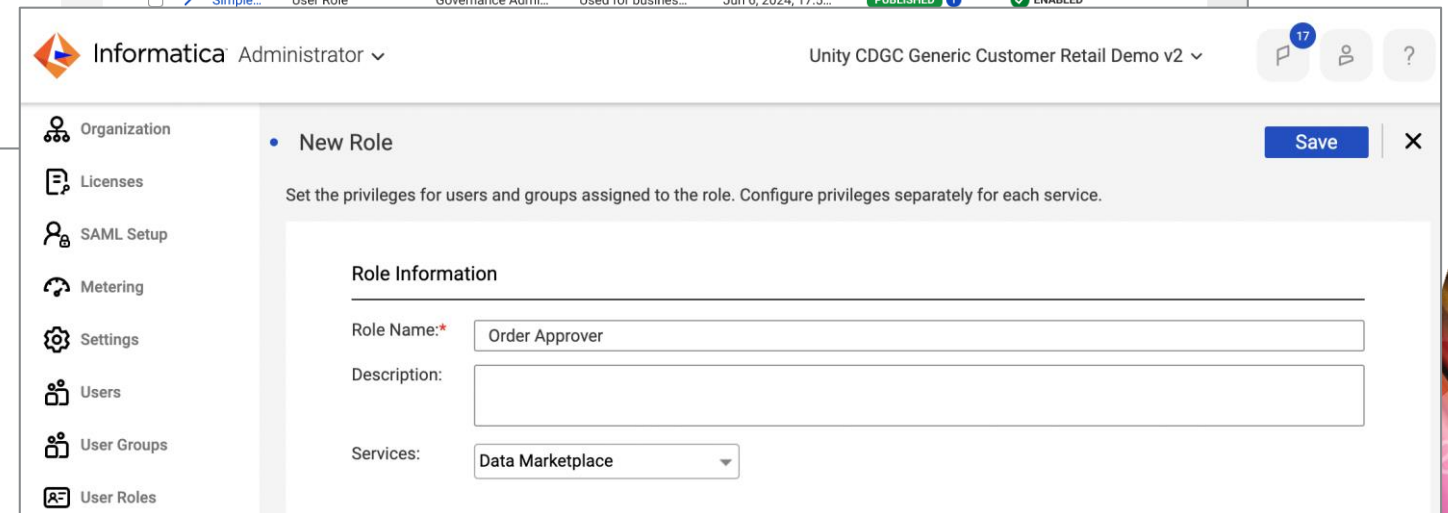
org switcher

**Access Control**

Access Policies Users

Policies (5)

<input type="checkbox"/>	Name	Type	User	Description	Last Updated	Lifecycle	Status	
<input type="checkbox"/>	> Base Policy	User Role	Governance Admi...	Performing admi...	Jun 6, 2024, 17:55	DRAFT	DISABLED	...
<input type="checkbox"/>	> Simple...	User Role	Governance Admi...	Used for busines...	Jun 6, 2024, 17:5...	PUBLISHED	ENABLED	
<input type="checkbox"/>	> Base Policy	User Role	Governance Admi...	Performing admi...	Jun 6, 2024, 17:55	PUBLISHED	DISABLED	
<input type="checkbox"/>	> Simple...	User Role	Governance Admi...	Used for busines...	Jun 6, 2024, 17:5...	PUBLISHED	ENABLED	



Informatica Administrator

Unity CDGC Generic Customer Retail Demo v2

Organization Licenses SAML Setup Metering Settings Users User Groups User Roles

**New Role** Save X

Set the privileges for users and groups assigned to the role. Configure privileges separately for each service.

**Role Information**

Role Name:\*

Description:

Services:



# Metadata Access Control

- As in CDGC, Metadata Access Control will be available in Marketplace
- The most popular use case is segmentation of information using asset groups (at the category level)

- Administrators can:

- Create user groups
- Create asset groups
- Link them together with policies
- Assign a category to an asset group (will propagate down hierarchy as part of sync job)

The image shows two screenshots from a user interface. The left screenshot is titled 'New Group' and contains a form for creating a group. It has a sub-header 'Group Information' and two input fields: 'Name: \*' with the value 'HR Leaders' and 'Description:' with the value 'HR Leadership Team'. A blue arrow points from this form to the right screenshot. The right screenshot is titled 'Super Secret HR' and is labeled 'ASSET GROUP'. It has a sub-header 'Asset Group Properties' and two fields: 'Name: \*' with the value 'Super Secret HR' and 'Description:' with the value 'Very sensitive human resources data, visible only to HR leaders'.

- This will mean users can:

- See all categories that do not have an asset group
- See any categories that are assigned an asset group that they have permissions for
- Not see categories (their contents, e.g. data collections) they are not in the asset group for.

- Mostly they will be hidden completely. Where necessary will see obfuscated names – XXXXX.

The image shows two screenshots from a user interface. The left screenshot is titled 'Super Secret HR' and is labeled 'USER GROUP POLICY'. It has a sub-header 'Condition' and a form with a dropdown menu 'Asset Type' set to 'Is Any'. Below this is an 'And' condition with a sub-condition 'Asset Groups' set to 'Is Any Of' and 'Super Secret HR'. There is a '+ Add Predicate' button. Below the condition is a 'Read' permission with a '+ Add Permission' button. A blue arrow points from this screenshot to the right screenshot. The right screenshot is titled 'Super Secret HR Data' and has an information icon. It contains the text: 'Details about pay, bonuses, schemes under consideration, appraisals, etc. Should only be visible to HR leaders.'



# Related Changes

- Browse category re-vamp
- Embedded experience for creating categories and data collections
- Stakeholder changes
  - Tabs on category and collection
  - Panels on summary page
  - Tile on orders, requests & accesses
  - New 'edit' experience with **additive** propagation (different from CDGC)
  - Owners column / filter > Stakeholders
- Scope of OOTB roles changing
  - Delivery Template Owner can **only** create / update formats, methods and templates
  - Technical Owners can edit **all** fields and relationships on Data Collection
  - IICS Admin will not be given full marketplace permissions as part of this role
  - All users can request withdrawal of consumer accesses (including consumer)

The image displays two overlapping screenshots of the Informatica Data Marketplace user interface. The top screenshot shows the 'EME - HR' category summary page. It features a 'Category Summary' table with fields: REFERENCE ID (CAT-9), NAME (EME - HR), DESCRIPTION (Category for the EMEA HR), STATUS (Active), PARENT CATEGORY (-), and ASSET GROUPS. To the right, there are three summary tiles: '1 Data Collection', '2 Sub-Categories', and '2 Data Collections in Sub-Categories'. Below these is a 'Stakeholders' section with a 'View All' link and a list of roles including 'CUSTOM DATA MARKETPLACE CATEGORY OWNER'. The bottom screenshot shows the 'New Category' creation form, which includes a 'New' button, a navigation menu (Home, Browse, Search, Compare, My Data), and input fields for 'REFERENCE ID' and 'NAME\*'. A 'Create' button is visible in the top right of the form. A separate 'Stakeholders' panel is shown in the bottom right, listing roles: 'CUSTOM DATA MARKETPLACE ADMINISTRATOR' (PK) and 'CUSTOM DATA MARKETPLACE DATA COLLECTION OWNER' (AP), with a 'View All' link.

# Custom Roles

- To allow for custom roles and metadata access control, we have formalized the 7 OOTB roles
  - These are still available with the tweaks mentioned
  - Their capabilities are controlled through:
    - Policies, e.g., Data Owner can create, read, update and delete categories
    - Privileges, e.g. Data Owner can 'Approve or Reject Orders'
- Customers can use these mechanisms to create their own custom roles by:
  - Creating a new role with a new name
  - Allocate privileges
  - Cloning or building policies (inc. manage access permission for 'stakeholder' roles)
  - Assigning stakeholders on assets in Marketplace

Informatica Administrator

Organization Licenses SAML Setup Metering Settings Users

**New Role**

Set the privileges for users and groups assigned to

**Role Information**

Role Name:\*

Description:

Data Marketplace Administrator

Set the privileges for users and groups assigned to the role. Configure privileges separately

Services:

**Assets** **Features**

**Feature Name**

Access Data Marketplace

Approve or Reject Orders

Cancel Your Orders And Data ...

Complete or Reject Data Coll...

Configure And Manage Data ...

Informatica Metadata Command Center

**Access Control**

Access Policies Users

**Policies (7)**

Name	Type	User	Description	Last Updated	Lifecycle	Status
> <input type="checkbox"/> CDMP Data Collecti...	User Role Policy	Custom Data Marketplace Data Collection Owner	CDMP Data Colle...	Oct 13, 2024, 11:47 P...	PUBLISHED	ENABLED
> <input type="checkbox"/> CDMP Category Ow...	Stakeholder Role Policy	Custom Data Marketplace Category Owner	CDMP Category ...	Oct 13, 2024, 5:29 AM	PUBLISHED	ENABLED
▼ <input type="checkbox"/> CDMP Data Collecti...	Stakeholder Role Policy	Custom Data Marketplace Data Collection Owner	CDMP Data Colle...	Oct 13, 2024, 5:29 AM	PUBLISHED	ENABLED

**Edit Stakeholders**

Select Roles: Custom Data Marketplace Data Collection Owner

Users (2) User Groups

NAME	EMAIL
<input checked="" type="checkbox"/> Ankit P	aphondani@informatica.com
<input checked="" type="checkbox"/> Kate Amory	kamory@informatica.com

**Stakeholders (3)**

STAKEHOLDER ROLE	STAKEHOLDER NAME
Custom Data Marketplace Administrator	PK Prasanna Kharche
Custom Data Marketplace Data Collection Owner	AP Ankit P
Custom Data Marketplace Data Collection Owner	KA Kate Amory

# Additional Features

- Ability to delete additional assets via API
  - Consumer accesses in withdrawn status
  - Orders without consumer accesses
  - Requests with any status
- Read APIs available on CDGC Gateway
- Additional asset types
  - Added 60 more asset types from CDGC
  - Inc. Looker, SAP BW, Tibco, Oracle BIEE
- Usability Improvements
  - Carousel shows new Data Collections
  - Certified use prioritized in check-out
  - No warning if no certified use listed
  - Link to CDGC asset in jumbo tooltip

Data Governance and Catalog Asset Types				
NAME ^	COMPATIBILITY ↕	STATUS ↕	DESCRIPTION	
Document Report	Available	<input type="checkbox"/> Inactive	MicroStrategy Document Report	...
ESRI Shape File	Available	<input type="checkbox"/> Inactive	TIBCO Spotfire ESRI Shape File	...
Excel Workbook	Available	<input type="checkbox"/> Inactive	TIBCO Spotfire Excel Workbook	...
Excel Worksheet	Available	<input type="checkbox"/> Inactive	TIBCO Spotfire Excel Worksheet	...
External Table	Available	<input type="checkbox"/> Inactive	Oracle Business Intelligence Enterprise ...	...
External Table	Available	<input type="checkbox"/> Inactive	An external table is a table whose data c...	...
External View	Available	<input type="checkbox"/> Inactive	SAP S4HANA External View	...
File	Available	<input type="checkbox"/> Inactive	A collection of data stored in one unit	...
File	Available	<input type="checkbox"/> Inactive	Qlik View File	...
File	Available	<input type="checkbox"/> Inactive	TIBCO Spotfire File	...

41 - 50 of 121 Object types      < 5 of 13 >      Rows: 10

### Recent Data Products

**Approved Providers** ⓘ

List of all the approved providers of...

ORDERS ⓘ      DELIVERY ⓘ

1 Ordered

**Marketplace (CDMP) ...** ⓘ

Provides a comprehensive summary of...

ORDERS ⓘ      DELIVERY ⓘ

1      2 Fulfilled

**Year To Year (YOY) Sal...** ⓘ

Analyze sales revenue over the last 3...

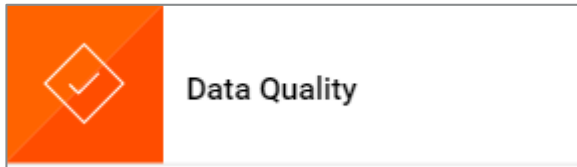
ORDERS ⓘ      DELIVERY ⓘ

84 Fulfilled

# Cloud Data Quality

# Cloud Data Quality

## What's New in Cloud Data Quality - November 2024?



- Data Observability
  - Allow user to choose per catalog source: algorithms to run algorithm sensibility
  - Email Notifications for events identified during the DO Job execution
  - Increased the number of supported data elements per catalog source to 200,000
- Verifier:
  - Consumption based license for Verifier data.
  - Add Use Abbreviation as an option to the Alias Handling
  - Add Australian Enhancement ports, Local Government Area Name and Code
- New Data Quality Bundle: India
  - Validation rules for India: PAN number, Aadhaar, EPIC/Voter ID, Passport, Driver license, and more.
- Enabled custom dimensions for Data Quality Assets
- Expose Precision/Scale Management to Advanced Rule User
- Exception Task:
  - Expanded PAM to match profiling PAM, including source mapplets
  - Limit number of exception rows per task to 20,000

# Profiling

## What's New in Profiling - November 2024?



### Data Profiling

- Profile Fixed Width files with flat file connection
- New Sources to profile::
  - PostgreSQL
  - Microsoft Fabric Data Warehouse
- Profiling Queries for Source Mapplets
- Usability:
  - Open Cloud Data Quality assets as a tab within Data Profiling service



### Data Governance and Catalog

- New Catalog Sources to profile::
  - IBM DB2 for LUW
  - DB2 for Z/OS
  - Hadoop HDFS
  - Microsoft Fabric OneLake
- Support to Key Vault

# Exception Management



# Exception Management

## Expanded sources to match profiling sources, including source mapplets

- Microsoft SQL and Azure Database
- Microsoft Azure Synapse SQL
- Oracle Database and Oracle Autonomous DB
- Microsoft SQL and Azure Database
- Microsoft Fabric Warehouse
- PostgreSQL (NEW)
- Amazon Redshift (using native V2 connector)
- Amazon Athena
- Snowflake (using native V2 connector)
- Google Big Query
- Delimited File on: File system | Amazon S3 | Azure ADLS Gen2 | Google Cloud Storage | Oracle Cloud Storage | Microsoft Fabric OneLake
- ODBC
- JDBC (using V2 connector)
- SAP ECC, SAP S/4 Hana and SAP BW
- Databricks Delta tables using IDMC connector
  - Secure agent
- Salesforce
  - Salesforce Sales Cloud, Salesforce Service Cloud
  - Salesforce Verticals: Salesforce Financial Cloud, Salesforce Health Cloud
  - Applications on Force.com
- Source mapplets
  - Amazon Redshift, Amazon S3, Azure Data Lake Store, Databrick, (ODBC), Flat File, Google BigQuery, Google Cloud Storage, Microsoft Azure, Synapse, Microsoft SQL Server, Oracle, Oracle (ODBC), Oracle Object Source, Salesforce, SAP BW, SAP Table, Snowflake. Athena, Microsoft Fabric OneLake , PostgreSQL

# Custom dimension

# Data Quality Assets

## Custom Dimensions

- **Use Case/User Statements**

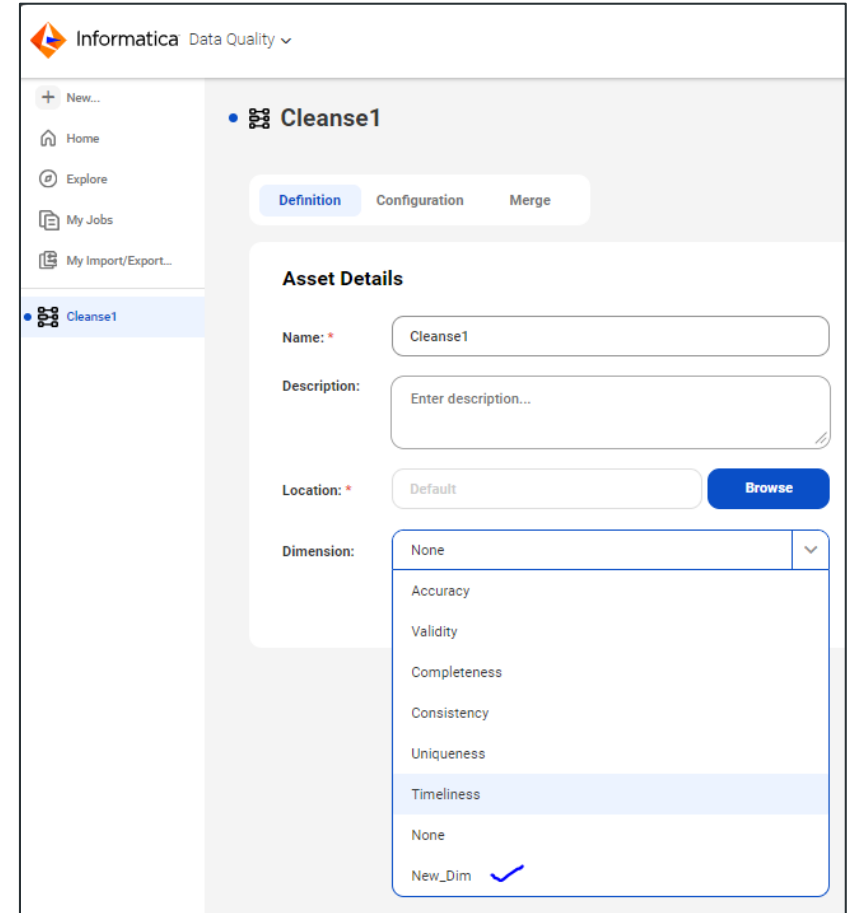
- Need to customize dimension and use in Data Quality Assets

- **Solution**

- Use Metadata Command Center (MCC) as the source of truth for dimensions
- Make available in Rule Spec, Cleanse, Parser, Verifier and Labeler

- **NOTES**

- The original 6 dimensions will ALWAYS show up, no matter if gets deleted in MCC
- Rename of a dimension being used in a CDQ asset will reflect in the DQ asset once page is refreshed.
- Rename of a dimension (which is part of 6 static DQ dimensions) will be treated as a new dimension
- If CDGC Dimensions fetch API call fail, a warning message is shown
- User can delete a dimension in MCC even if it is used in a DQ Asset
  - By marking DQ asset dirty & default dimension will be Accuracy, User needs to manually update the CDQ Asset and choose an existing dimension



The screenshot displays the Informatica Data Quality interface for configuring a 'Cleanse1' asset. The interface includes a navigation sidebar on the left with options like 'New...', 'Home', 'Explore', 'My Jobs', and 'My Import/Export...'. The main content area shows the 'Cleanse1' asset configuration with tabs for 'Definition', 'Configuration', and 'Merge'. The 'Asset Details' section contains the following fields:

- Name:** Cleanse1
- Description:** Enter description...
- Location:** Default (with a 'Browse' button)
- Dimension:** A dropdown menu is open, showing a list of dimensions: None, Accuracy, Validity, Completeness, Consistency, Uniqueness, Timeliness, None, and New\_Dim (which is selected with a checkmark).

# Data Quality Assets

## Custom Dimensions

- Management to Dimensions is performed in Metadata Command Center (MCC):
  - Path: MCC->Customize->
  - When DELETING a dimension, no checks are done against Data Quality Assets
  - RENAMING the fixed 6 dimensions will be considered as a NEW dimension in Data Quality Assets
- Scorecards with new dimensions is not supported – **COMING SOON**

The screenshot shows the 'Customize' interface with three tabs: 'Asset Customization', 'Custom Catalog Source Types', and 'Metadata Models'. The 'Asset Types' section is active, displaying a search bar and a list of asset types under 'Business Assets'. The 'Details' section shows 'View all attributes that are associated with asset types' and a 'Create' button. Below this is a table of 'Attributes (23)' with columns for Name, Data Type, Asset Type, Mandatory, and Search and Filter.

Name	Data Type	Asset Type	Mandatory	Search and Filter
<input type="checkbox"/> AI Model Purpose	Text Field - Multi Line	AI Model	No	Yes
<input type="checkbox"/> Architecture Type	Dropdown - Single Select	AI Model	No	Yes
<input type="checkbox"/> Asset ID	Text Field - Single Line	System	No	Yes
<input type="checkbox"/> Business Logic	Text Field - Multi Line	Metric +2	No	No
<input type="checkbox"/> Dim	Dropdown - Single Select	Data Quality Rule Occu... +1	Yes	Yes
<input type="checkbox"/> Dimension	Dropdown - Single Select	Data Quality Rule Tem... +1	Yes	...
<input type="checkbox"/> Effective Date	Date	Policy	No	No
<input type="checkbox"/> End Date	Date	Policy +1	No	No

The screenshot shows the 'Dimension' configuration page. At the top right, there are buttons for 'Save', 'Publish', and 'X', along with lifecycle information: 'LIFECYCLE DRAFT', 'LAST UPDATED Oct 21, 2024, 8:04 AM', and 'LAST UPDATED BY Paul O'Shea TEAM'. The 'Attribute Properties' section includes fields for Name, Data Type, Subtype, and Values. The 'Values' field is a list of radio buttons with corresponding values.

Default	Value
<input checked="" type="radio"/>	cdgc_dim
<input type="radio"/>	TEST_MO_DQ_rename1
<input type="radio"/>	Cleanse_P1_Edited
<input type="radio"/>	Cleanse_P_
<input type="radio"/>	Adv_rule_p0_test
<input type="radio"/>	uNIQvEnESs
<input type="radio"/>	AAAAAAAAcçÉÉÉÉlll

# Precision/Scale Management to Advanced Rule

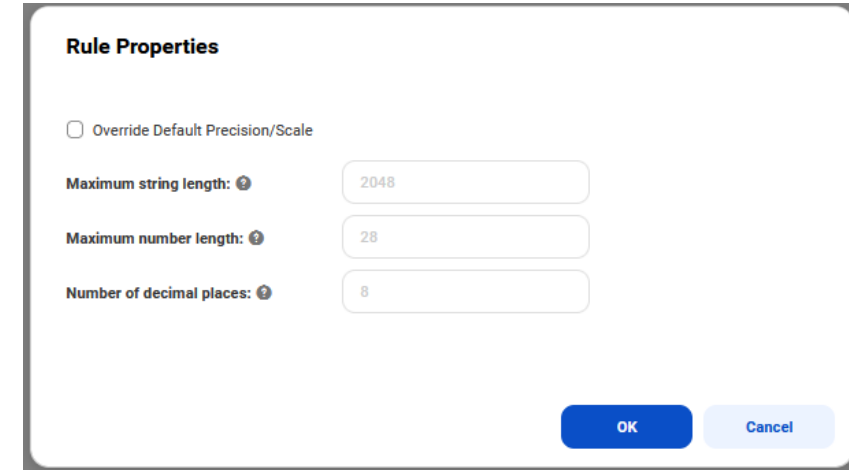
# Data Quality Assets

## Precision/Scale Management to Advanced Rule

- Rule properties tab add in **Advanced Rule**.

Default precision values:

- String – 2048
  - Number – 28
  - Decimal Scale – 8.
- User can override the default value
    - Greater the precision, more memory consumed on agent



**Rule Properties**


Override Default Precision/Scale

Maximum string length: 2048

Maximum number length: 28

Number of decimal places: 8

OK Cancel



**Rule Properties**

Override Default Precision/Scale

Maximum string length: 2048

Maximum number length: 28

Number of decimal places: 8

OK Cancel

# Verifier

# Verifier

## Consumption based license for Verifier data

- **Use Case**

- Address Verification for 240+ localities in **batch mode**

- **Solution**

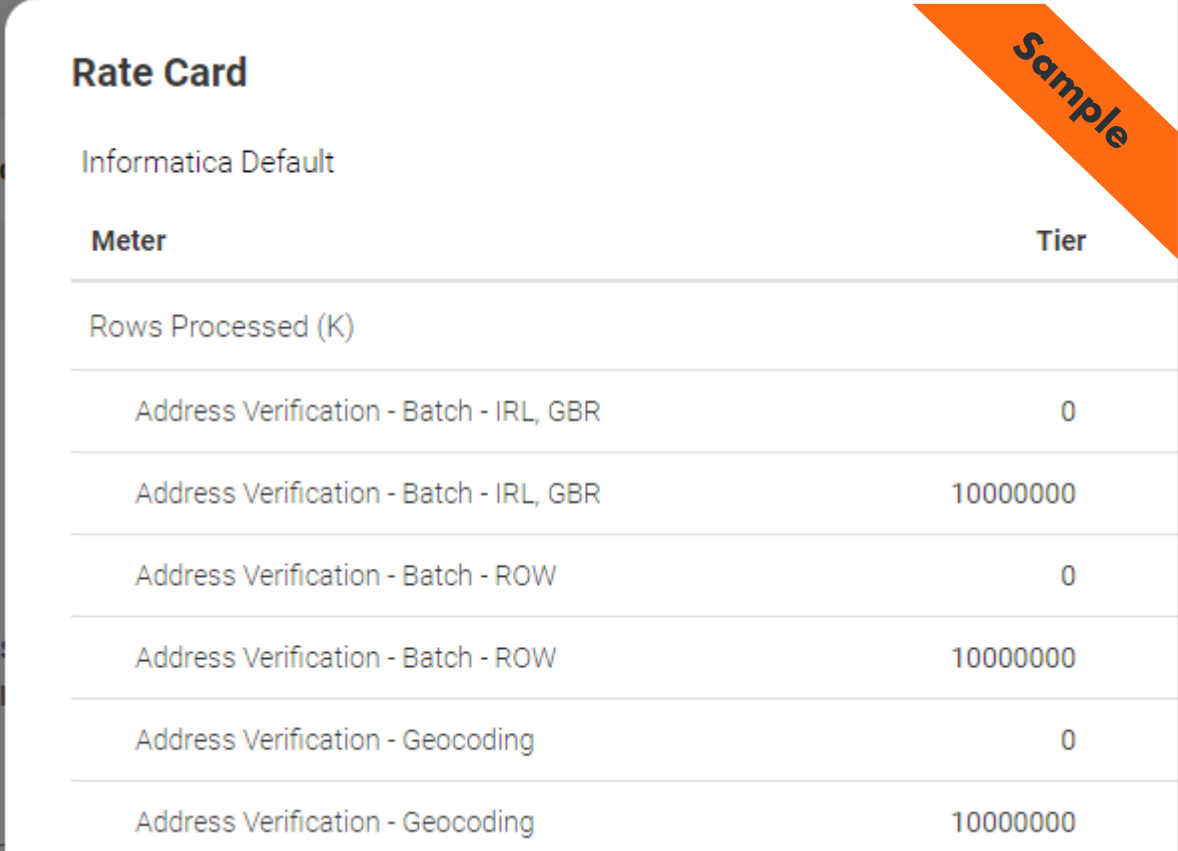
- Verifier asset available to use in CDI mappings (since 2020), but licensed in consumption model (IPU)

- **Benefits**

- Added functionality using existing IPU pool
- Experience remains the same
  - Verifier asset in Data Quality Service
  - Verifier transformation in Cloud Mapping Designer

- **NOTES:**

- Realtime mode continues to requires a DaaS license
- MDM continues to require license to address data



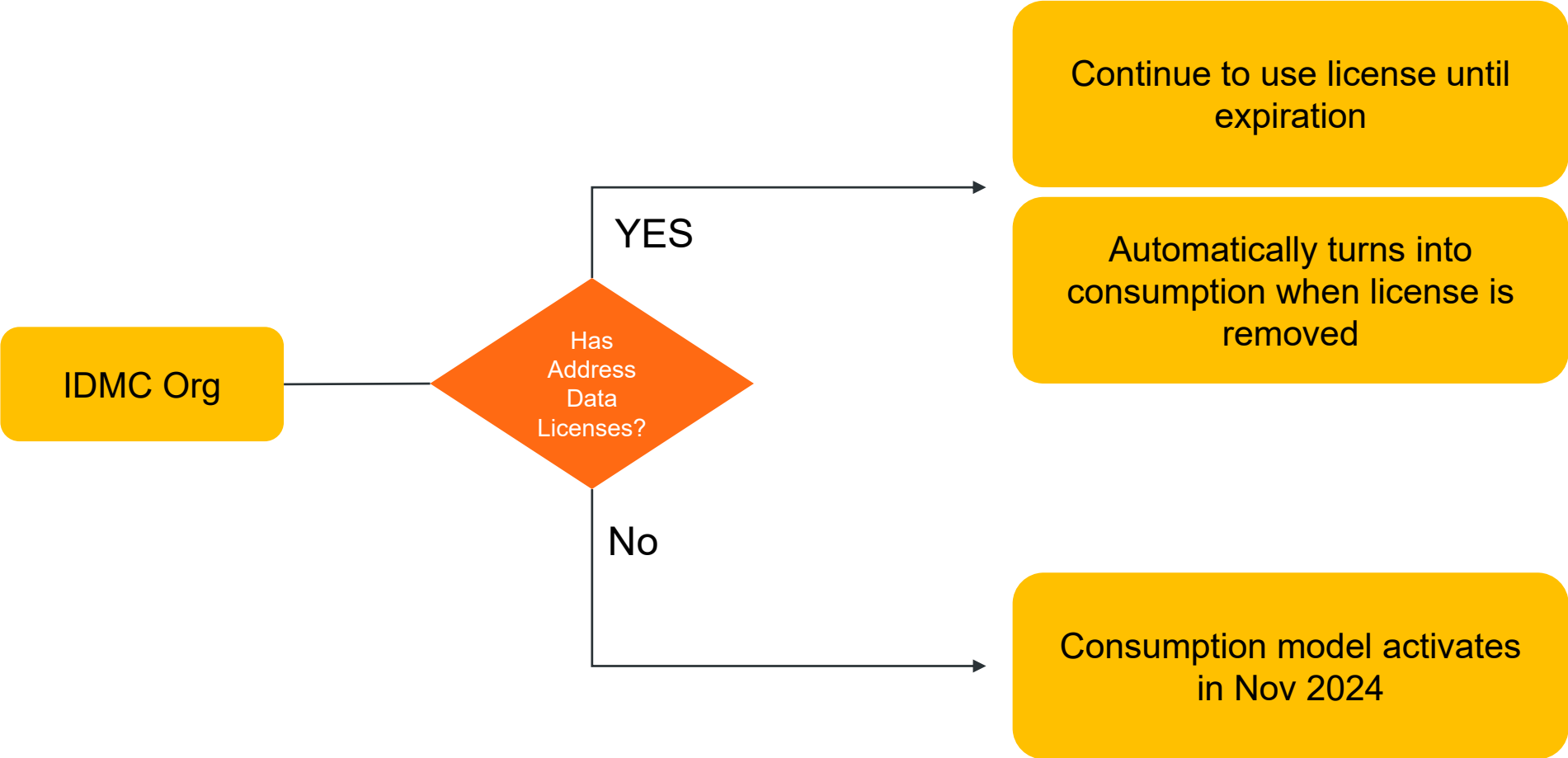
A sample rate card table for Informatica Default. The table has two columns: 'Meter' and 'Tier'. The 'Meter' column lists various address verification and geocoding tasks, and the 'Tier' column shows the corresponding values. An orange diagonal banner with the word 'Sample' is in the top right corner of the table area.

Meter	Tier
Rows Processed (K)	
Address Verification - Batch - IRL, GBR	0
Address Verification - Batch - IRL, GBR	10000000
Address Verification - Batch - ROW	0
Address Verification - Batch - ROW	10000000
Address Verification - Geocoding	0
Address Verification - Geocoding	10000000



# Verifier

Transition from old model to new consumption model for existing customers



# Verifier

## Use Abbreviation as an option to the Alias Handling

- Actual verbiage for UseAbbreviation option in Alias Handling drop down is Abbreviation.
- Returns the address element in its abbreviated form if the reference data contains the abbreviation.

Accepts a string that can contain a single address element or multiple elements. Connect address fields to the common hybrid format combines house number, street name, and apartment information in a single field and...

**Properties** ✕

---

**Input**    **Process**    **Result**    **Standardization**

Language: ⓘ   

Script: ⓘ   

Descriptor standardization: ⓘ   

Casing: ⓘ   

Alias handling: ⓘ

Full

Full

Preserve official

Postal admin

Abbreviation

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Street_1	Locality_1	Country_Name_1	Address_Lines_1	Formatted_Address_Lines_1	Verification_Status_Code	Result_Quality	Match_Percentage	Address_Type	Street	HouseNumber	PostalCode	Locality	Country
2	APPIN PL	ST MARYS	AUSTRALIA	3 APPIN PL	3 APPIN PL	C	4	92.92 S	APPIN PLACE	3	2770	DUNHEVED	AUSTRALIA	
3	290TH ST	DALLAS	UNITED STATES	32900 290TH ST	32900 290TH ST	A	6	99.99 S	290TH ST	32900	57529	Dixon	USA	
4														
5														

With Abbreviation

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Street_1	Locality_1	Country_Name_1	Address_Lines_1	Formatted_Address_Lines_1	Verificatio	Result_Qu	Match_Percentage	Address_Type	Street	HouseNumber	PostalCode	Locality	Country
2	APPIN PLACE	DUNHEVED	AUSTRALIA	3 APPIN PLACE	3 APPIN PLACE	C	4	92.92 S	APPIN PLACE	3	2770	DUNHEVED	AUSTRALIA	
3	290TH ST	DIXON	UNITED STATES	32900 290TH ST	32900 290TH ST	A	6	99.99 S	290TH ST	32900	57529	Dixon	USA	
4														
5														

Without Abbreviation



# Verifier

Two new ports added to Enrichments – Country Specific – Australia

- **Local Government Area Name:** The name of the Local Government Area (LGA) in which an address is located.
- **Local Government Area Code:** A five-digit numeric code that identifies the Local Government Area (LGA) in which an address is located.

Australia

- Geocoded National Address File Identifier
- Level One Statistical Area 11-Digit
- Level One Statistical Area 7-Digit
- Level Two Statistical Area 9-Digit
- Level Two Statistical Area 5-Digit
- Level Two Statistical Area Name
- Level Three Statistical Area 5-Digit
- Level Three Statistical Area Name
- Level Four Statistical Area 3 Digit
- Level Four Statistical Area Name
- Greater Capital City Statistical Area 5-Char
- Greater Capital City Statistical Area Name
- State or Territory Name
- State or Territory Code
- Census Collection District Code 2006
- Mesh Block 11-Digit 2011
- Mesh Block 11-Digit 2016
- Local Government Area Name
- Local Government Area Code

	T	U
	Local_Government_Area_Name	Local_Government_Area_Code
	Cumberland	12380
	Cumberland	12380
	Swan	58050
	Cumberland	12380
	Swan	58050
	Cumberland	12380
	Cumberland	12380

# Data Observability

# Data Observability

## Email notifications

- Enable Data Observability notifications in new setting
- Event Email will enable Data Observability Emails sent after Data Observability job execution

**Notification Settings**

Select categories of notifications that you'd like to receive by email or email summary.

**Notification Category**

	Subcategory	Event Email	Email Summary
Collaboration			
Ticket and workflow			
Change Notification			
<b>Data Quality</b>	Data Quality, Failure, Anomaly	<input type="checkbox"/>	<input type="checkbox"/>
	Data Observability	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
CLAIRE Recommendations			
Admin Notification			
User Jobs			

**Email Preferences**

Send Summary Email: 12 Hours

# Data Observability

## Email notifications

- Sent to Stakeholders of assets with events
- One Email per Stakeholder
- Content:
  - Summary of all events created during the last job execution
  - List of affected assets (up to 10)

Data Observability Job Summary for FS1

admin@informaticacloud.com  
To: Choudhury, Hena  
Mon 28 Oct 2024 18:47

**Informatica** Organisation: 170ct2024\_DQDO\_Dev\_01  
Date: Oct 28, 2024, 5:47 PM

Hi,  
This email summarizes a Data Observability task job that ran in the above organization.

**Job Summary: Data Observability Task has identified '16' events.**

**Data Observability Job:** FS1\_File System\_sync  
**Start Date:** Oct 28, 2024, 5:31 PM  
**End Date:** Oct 28, 2024, 5:47 PM  
**Started By:** 170ct2024\_DQDO\_Dev\_01  
**Scanned Assets:** 30  
**Detected Events:** 16

**New Events By Severity**

- High: 9
- Normal: 7

**New Events By Metric**

- Nulls: 4
- Distinct: 4
- Non-Distinct: 4
- Min Length: 2
- Row Count: 2

**Affected Assets**

Assets	Asset Type	Number of Events
<a href="#">kusuma.csv</a>	FlatFile	8
<a href="#">testBook.csv</a>	FlatFile	7
<a href="#">Email</a>	FlatField	4
<a href="#">Column1</a>	FlatField	4
<a href="#">SSN</a>	FlatField	3
<a href="#">column4</a>	FlatField	3
<a href="#">All_Names_Dev.csv</a>	FlatFile	1

# Data Observability

More granular configuration options

- Select metrics
- Select sensitivity
- Select algorithms uses
- Configure training phase (runs before Data Observability starts creating events)
- Either include list of metrics or exclude metrics

### Data Observability

To enable Data Observability you must enable Data Profiling.

Data Observability

---

#### Parameters

Minimum Number of Data Points:

---

#### Filters

Metric Filters:

No Filters  Filter Conditions

# Cloud Data Profiling



# Cloud Data Profiling

## What sources can you profile? (starting October 2024)

### Databases and applications:

- Microsoft SQL and Azure Database with Active Directory Authentication
- Microsoft Azure Synapse SQL (using Native driver)
  - Azure ADLS gen 2 and Azure Blob storage type
- Microsoft SQL and Azure Database
- Microsoft Fabric Warehouse **(NEW)**
- Oracle Database and Oracle Autonomous DB
- PostgreSQL **(NEW)**
- Amazon Redshift (using native V2 connector)
- Amazon Athena
- Snowflake (using native V2 connector)
- Google Big Query
- ODBC
- JDBC (using V2 connector)
- SAP ECC, SAP S/4 Hana and SAP BW
- Databricks Delta tables using IDMC connector
  - Secure agent
  - Advanced / Elastic

### • Salesforce

- Salesforce Sales Cloud, Salesforce Service Cloud
- Salesforce Verticals: Salesforce Financial Cloud, Salesforce Health Cloud
- Applications on Force.com

### File based sources:

- Delimited File on:
  - File system ,Windows Share/CIFS | Amazon S3 | Azure ADLS Gen2 | Google Cloud Storage | Oracle Cloud Storage | Microsoft Fabric OneLake
- PARQUET, AVRO and JSON files on:
  - Amazon S3 - [Requires Advanced / Elastic]
  - Azure ADLS Gen2 - [Requires Advanced / Elastic]
- PARQUET on Microsoft Fabric OneLake - [Requires Advanced / Elastic]

### Source mapplets

- Amazon Redshift, Amazon S3, Azure Data Lake Store, Databrick, (ODBC), Flat File, Google BigQuery, Google Cloud Storage, Microsoft Azure, Synapse, Microsoft SQL Server, Oracle, Oracle (ODBC), Oracle Object Source, Salesforce, SAP BW, SAP Table, Snowflake. Athena, Microsoft Fabric OneLake , PostgreSQL **(NEW)**

# Profiling of fixed width files

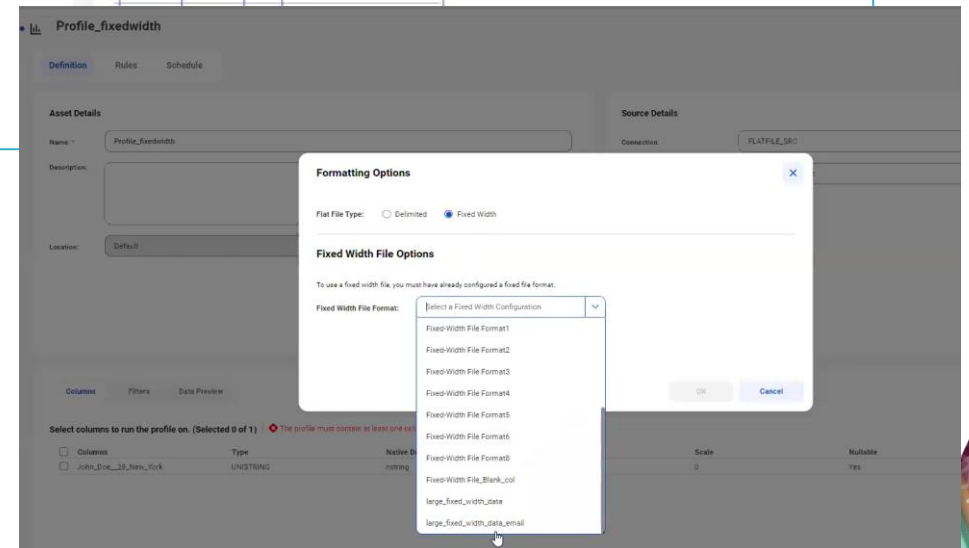
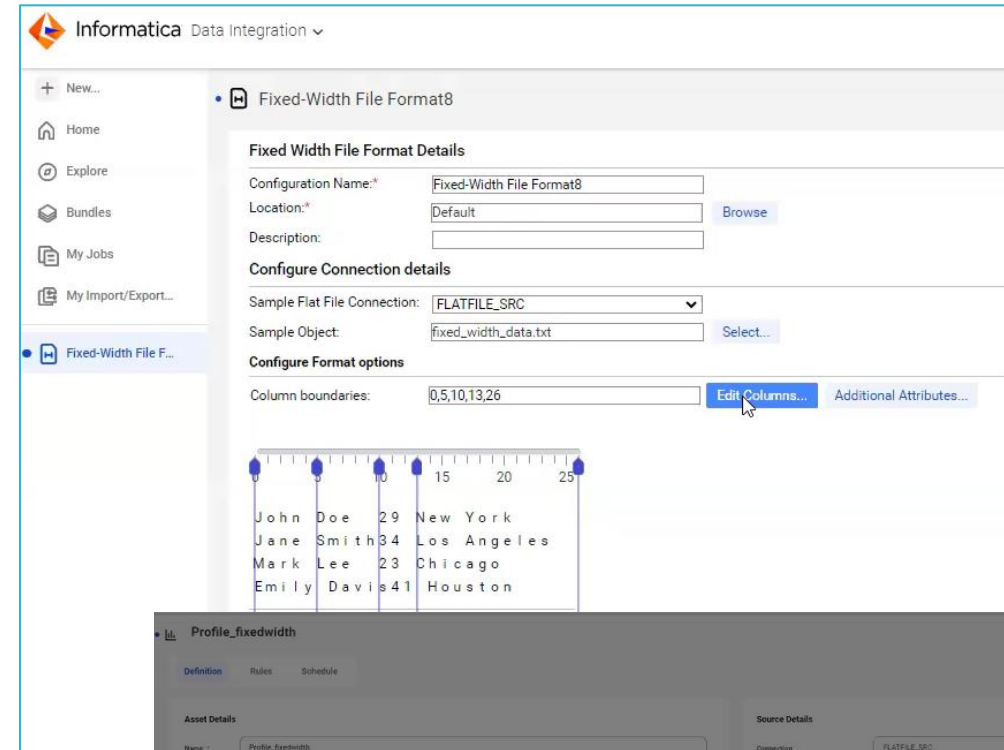
## Using Flat File connection

- **Use Case/User Statements**

- Profile Fixed Width files: a plain text file where each field value is the same width and padded with spaces.

- **Solution**

- PRE-REQUISITE:
  - In Data Integration service, configure “Fixed Width File Format”
- In Profiling Task:
  - Use “Flat File” connection and choose the file to profile
  - Configure profiling task with formatting option = “Fixed Width”
  - Choose the “Fixed Width File Format”
  - Preview the data to ensure the parser is formatting the data correctly
  - Run Profiling



# Profiling in Cloud Governance and Catalog

# Data Profiling in Data Governance and Catalog

What sources can you profile? (starting November 2024)

## Databases and applications:

- Microsoft SQL
- Microsoft Azure Synapse
- Microsoft Fabric Warehouse, Lakehouse
- Amazon AWS Athena, Redshift (using V2 connector)
- Oracle Database and Oracle Autonomous DB
- IBM DB2 for LUW **(NEW)**, DB2 for Z/OS **(NEW)**
- PostgreSQL
- My SQL
- Snowflake (using V2 connector)
- Google Big Query
- JDBC (using V2 connector)
- Maria DB
- Teradata
- Hive

- SAP ECC, SAP S/4 Hana, SAP BW, SAP HANA DB
- Databricks Managed tables and External tables in Delta Format
- Salesforce
  - Salesforce Sales Cloud, Salesforce Service Cloud
  - Salesforce Verticals: Salesforce Financial Cloud, Salesforce Health Cloud
  - Applications on Force.com

## File based sources:

- Delimited File on:
  - File system, Windows Share/CIFS | Amazon S3 | Azure ADLS Gen2 | Google Cloud Storage | Microsoft Fabric OneLake | Hadoop HDFS **(NEW)**
- PARQUET, AVRO and JSON files on:
  - Amazon S3 - [Requires Advanced / Elastic]
  - Azure ADLS Gen2 - [Requires Advanced / Elastic]
  - Google Cloud Storage - - [Requires Advanced / Elastic]
  - Microsoft Fabric OneLake - [Requires Advanced / Elastic] - **(NEW)**
  - Hadoop HDFS - **(NEW)**

# Data Profiling in Data Governance and Catalog

## New features

- Support for **Key Vault**
- Support profiling tables with **large number of columns** (using column partition)
- Profiling has been enhanced for the following sources:
  - Salesforce – added support to datatypes “**Rich Text Area**” and “**Multi pick list**”
  - Amazon Redshift
    - Support to profile **External Tables with “String” Data types**
      - Previous behavior: table is not profiled
      - New behavior: table is profiled, but columns with STRNG data type will have no statistics
    - Support to profile tables with **datatype “Super”**
      - Previous behavior: table is not profiled
      - New behavior: table is profiled.
        - If column Super can be mapped to a simple data type, profiling statistics will appear
        - If column Super contains semi-structure/hierarchical data, no profiling statistics will be presented

# Where data & AI come to

