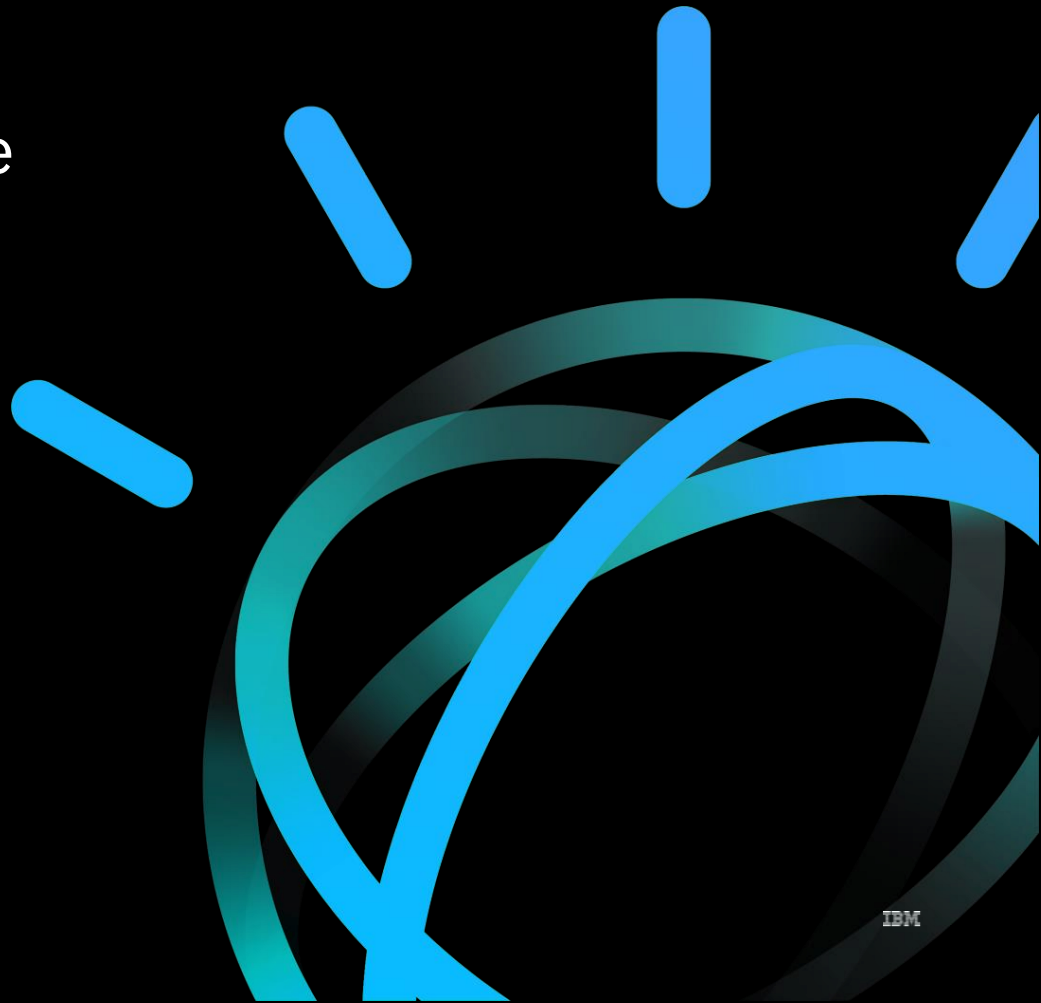
Enterprise Data Governance

IBM Knowledge Catalog

powered by Cloud Pak for Data



Mutiara Prahastuti
Data & AI Technical Specialist
IBM Indonesia
Mutiara.Prahastuti@ibm.com

IBM

Agenda

IBM Knowledge Catalog Overview

Delivering Business Ready Data

- **Step I – Catalog Foundation.**
- **Step II – Metadata Import and Discovery.**
- **Step III – Delivery of Business-Ready Data.**

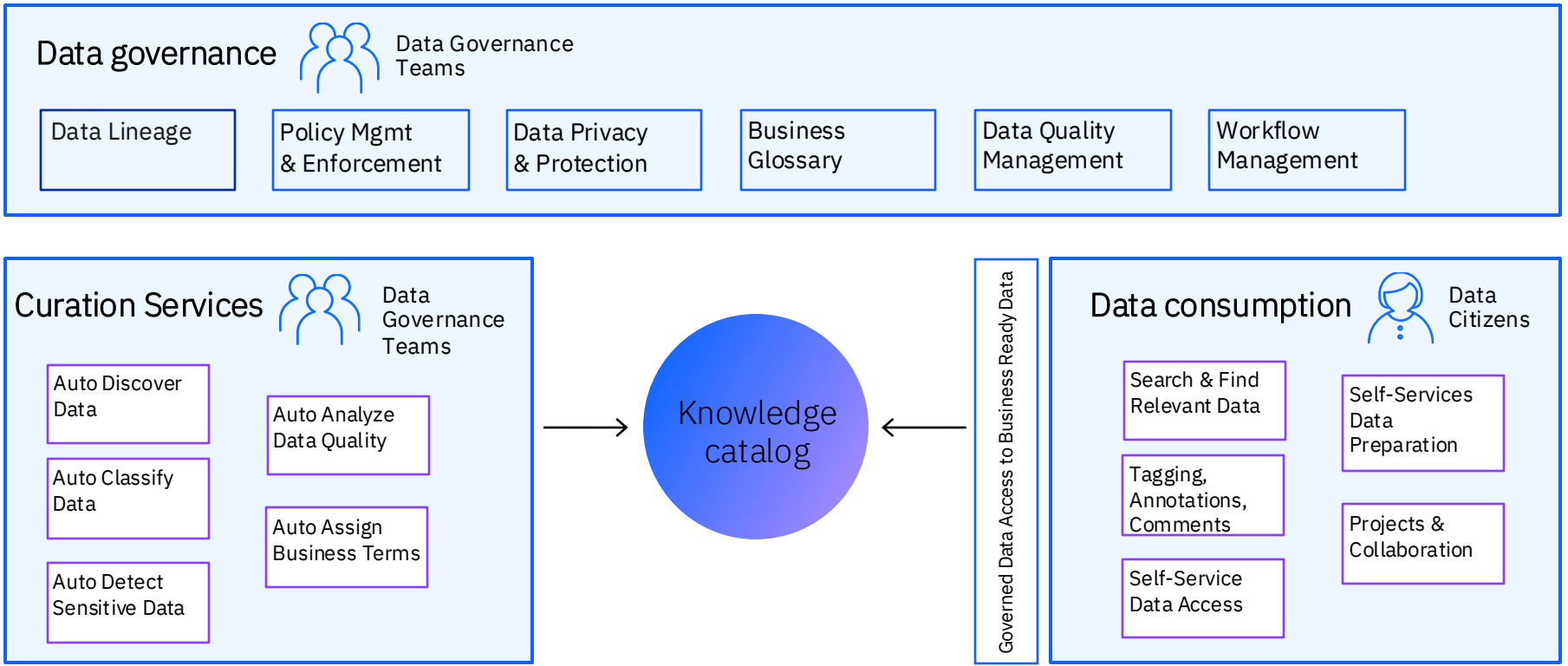
Data Fabric enables a hybrid multi-cloud data architecture

*Data Fabric connects the **right data**, at the **right time**, to the **right people**, from anywhere it's needed, with appropriate governance, leveraging current investments and future proofing new investments*

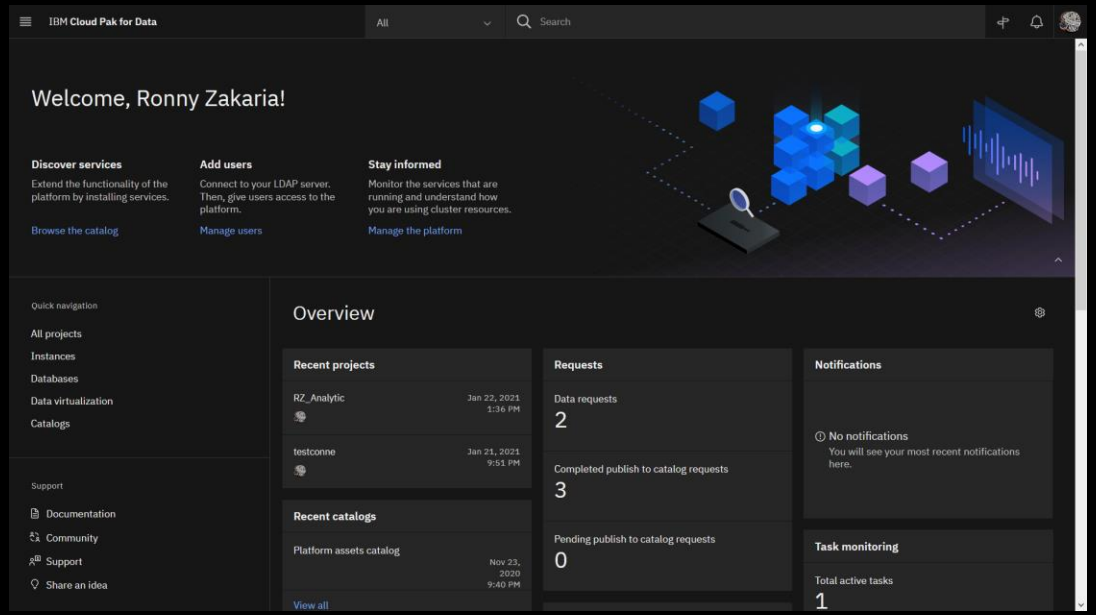
Ideally, all the data should be **well organised** and can be found easily

IBM Knowledge Catalog

Discover, trust and protect data with an automated metadata-driven data foundation



IBM Knowledge Catalog



Data Citizen – interested in consuming curated data to drive new Analytics

Information Analyst – interested in managing and curating business data

Data Steward – responsible for loading, managing and curating data across the Catalog

Quality Analyst – responsible for defining and managing Quality workspace, dimension and discovery

Agenda

IBM Knowledge Catalog Overview

Delivering Business Ready Data

- **Step I – Catalog Foundation.**
- **Step II – Metadata Import and Discovery.**
- **Step III – Delivery of Business-Ready Data.**

Summary & Discussion

IBM Knowledge Catalog

Delivering Business Ready Data to the Enterprise

Step I – Catalog Foundation

Configure Workflow and Business Process → Capture Business Language → Define Requirements & Responsibility → Define Domain and Ownership → Define Classification & Dimension

Catalog Foundation

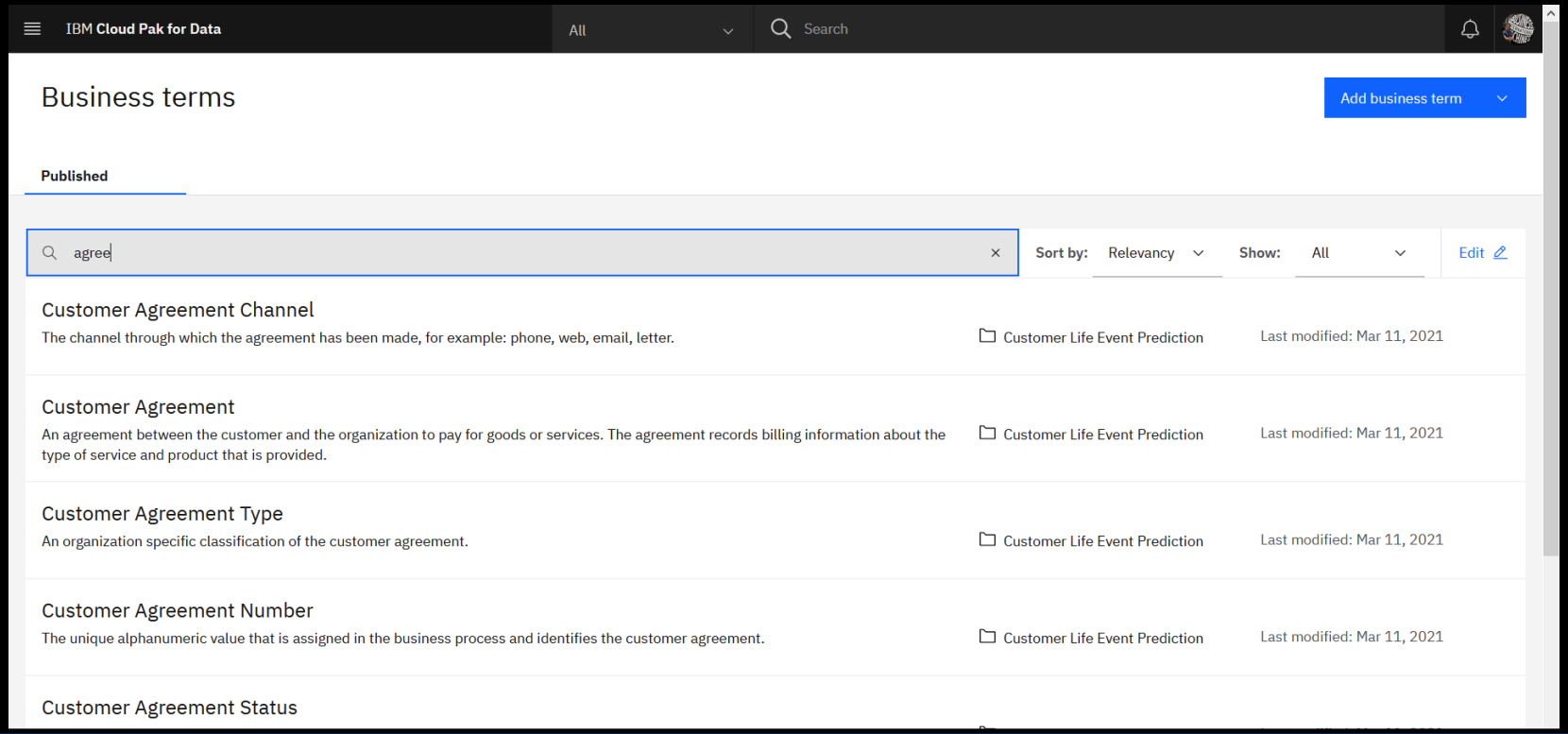
Glossary | Policies | Rules | Data Classes | Reference Data

IBM Knowledge Catalog

Data Catalog Platform

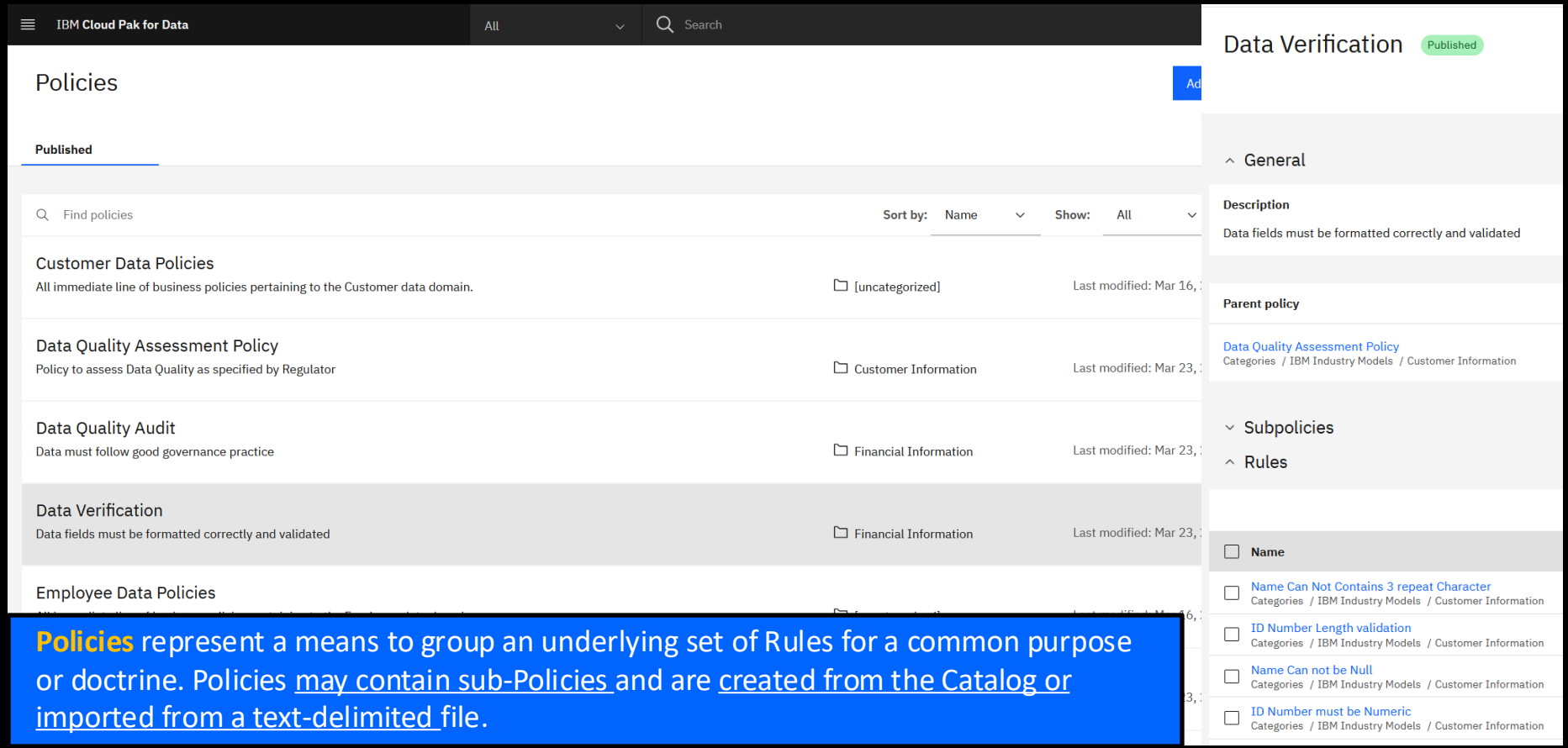
The first step is to build a solid foundation to support an Enterprise Data Catalog and enable data citizens to search and explore business-ready data. Terms apply meaning; Rules apply responsibility; Classifications side identification.

IBM Knowledge Catalog – Terms



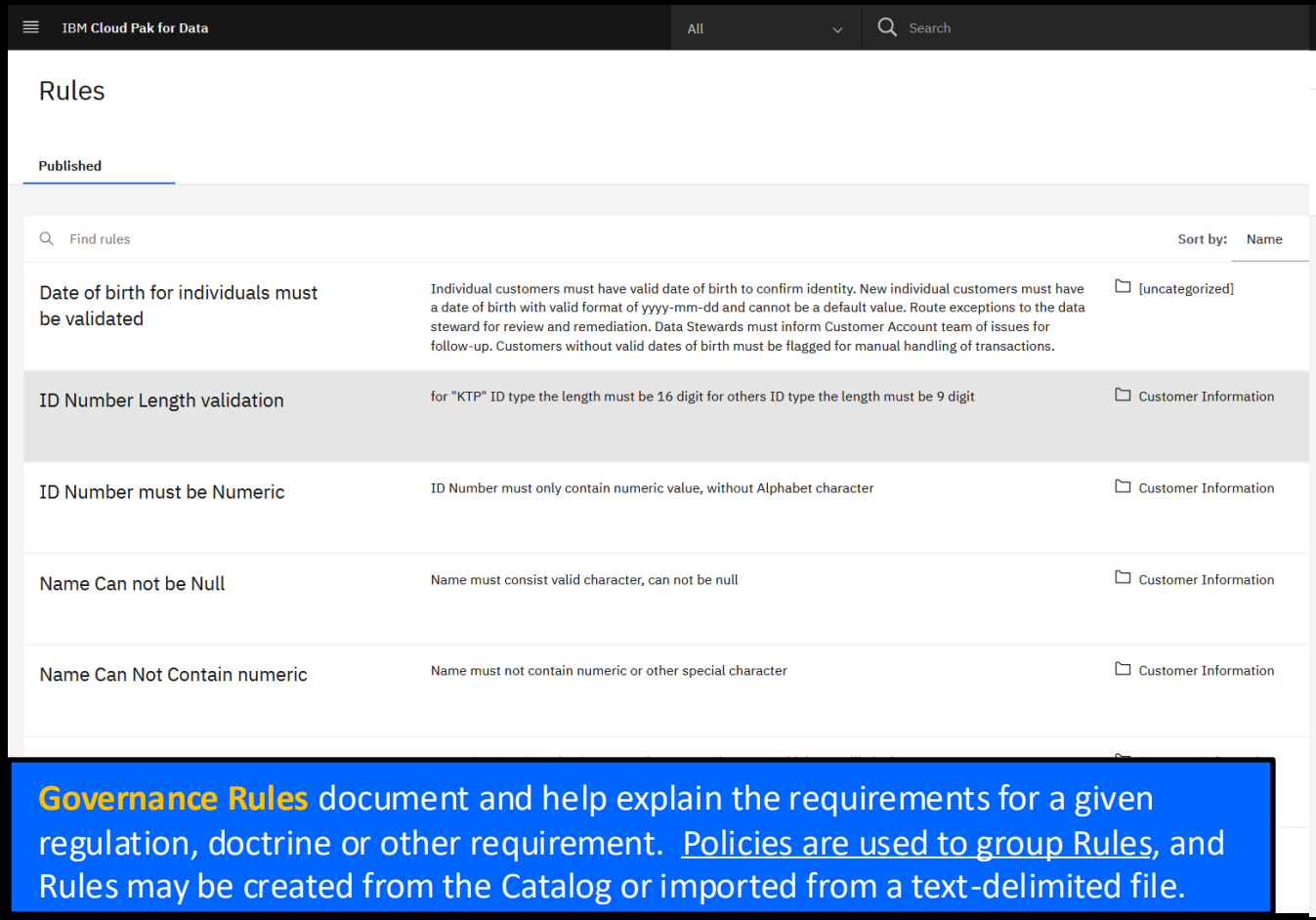
Glossary Terms may be created from the Catalog, or **may be imported from a text-delimited** file.
Glossary Terms are further **managed thru Workflow** which may require further reviewal and approval processes.

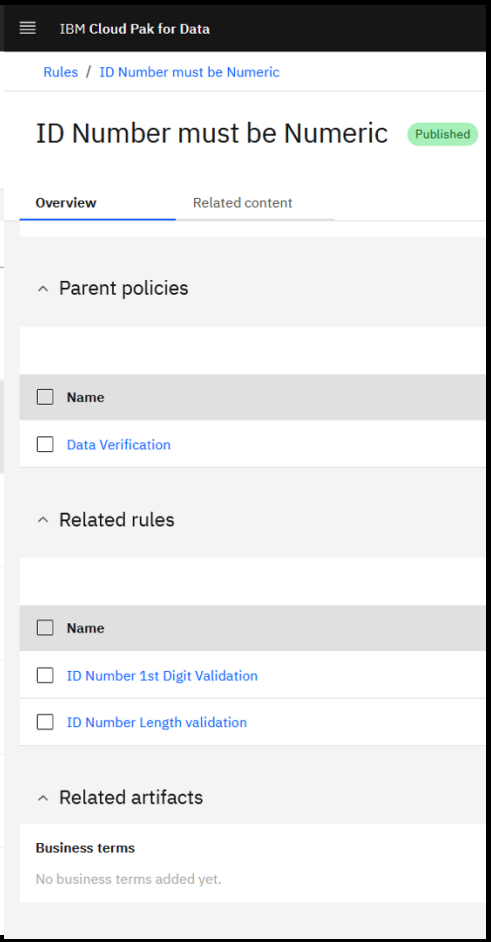
IBM Knowledge Catalog – Policies



Policies represent a means to group an underlying set of Rules for a common purpose or doctrine. Policies may contain sub-Policies and are created from the Catalog or imported from a text-delimited file.

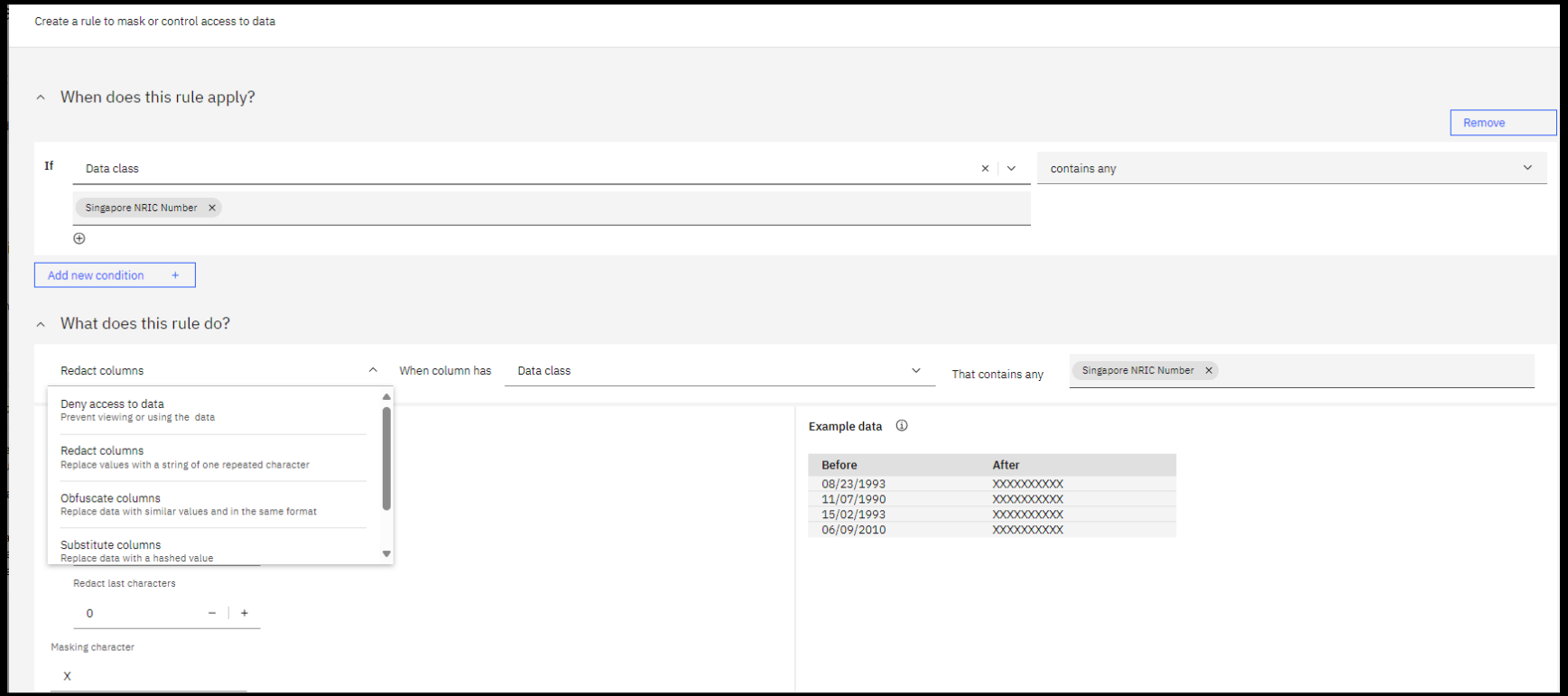
IBM Knowledge Catalog – Governance Rules





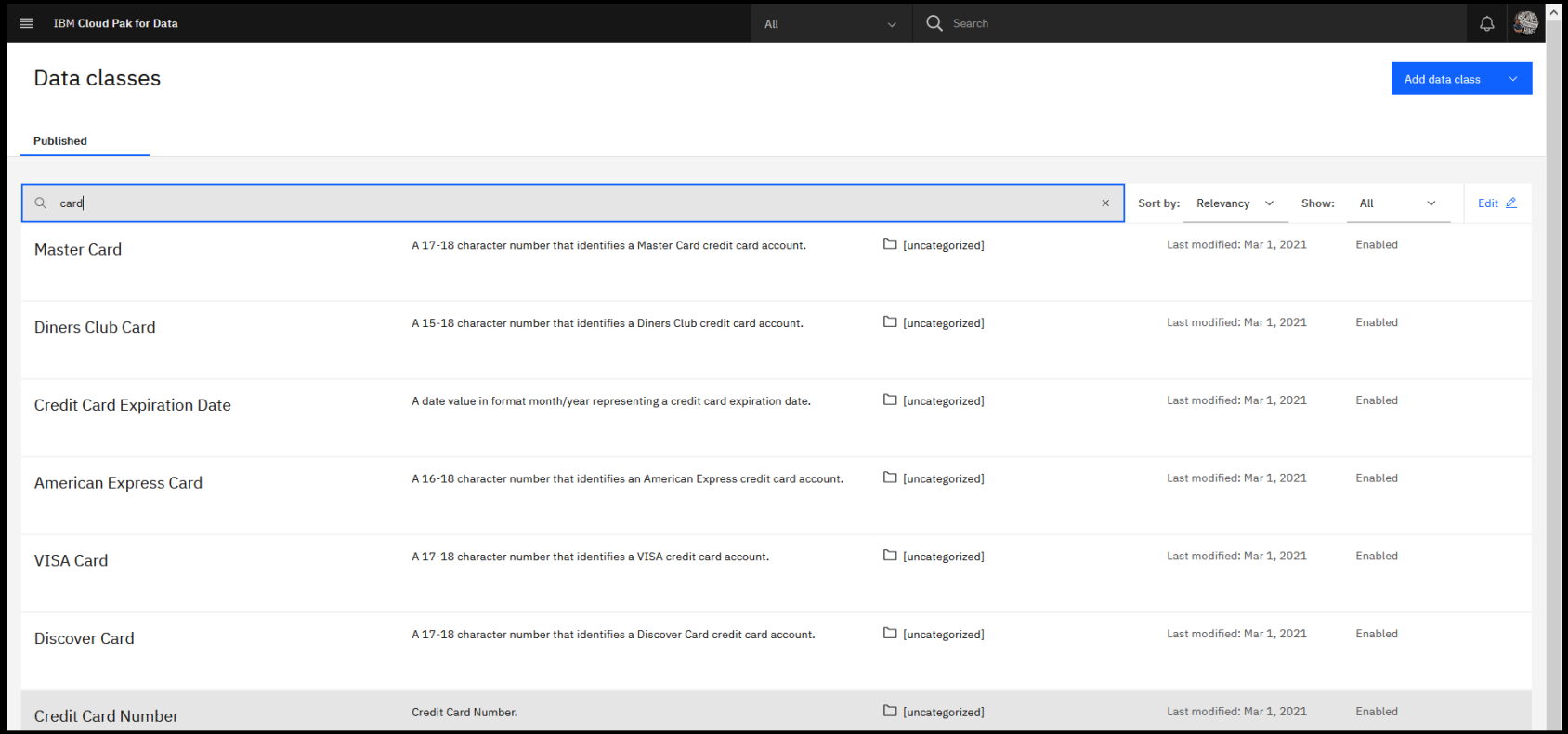
Governance Rules document and help explain the requirements for a given regulation, doctrine or other requirement. Policies are used to group Rules, and Rules may be created from the Catalog or imported from a text-delimited file.

IBM Knowledge Catalog – Data Protection Rules



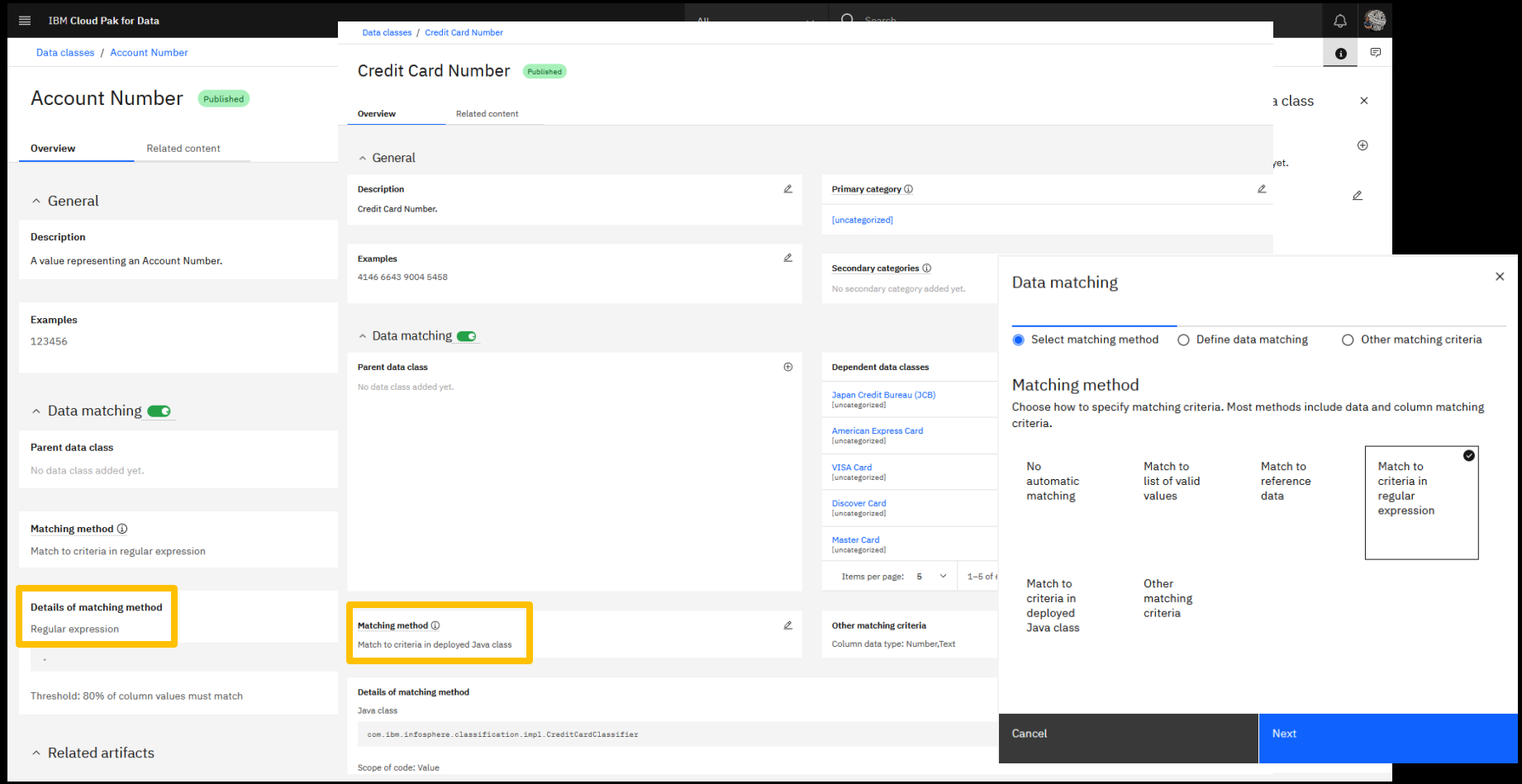
Data Protection Rules represent a set of instructions to deny access, redact, masking or filter data from a consumer (for purposes of Preview or Refinement) based upon the classification or other characteristic of the data.

IBM Knowledge Catalog – Data Classes

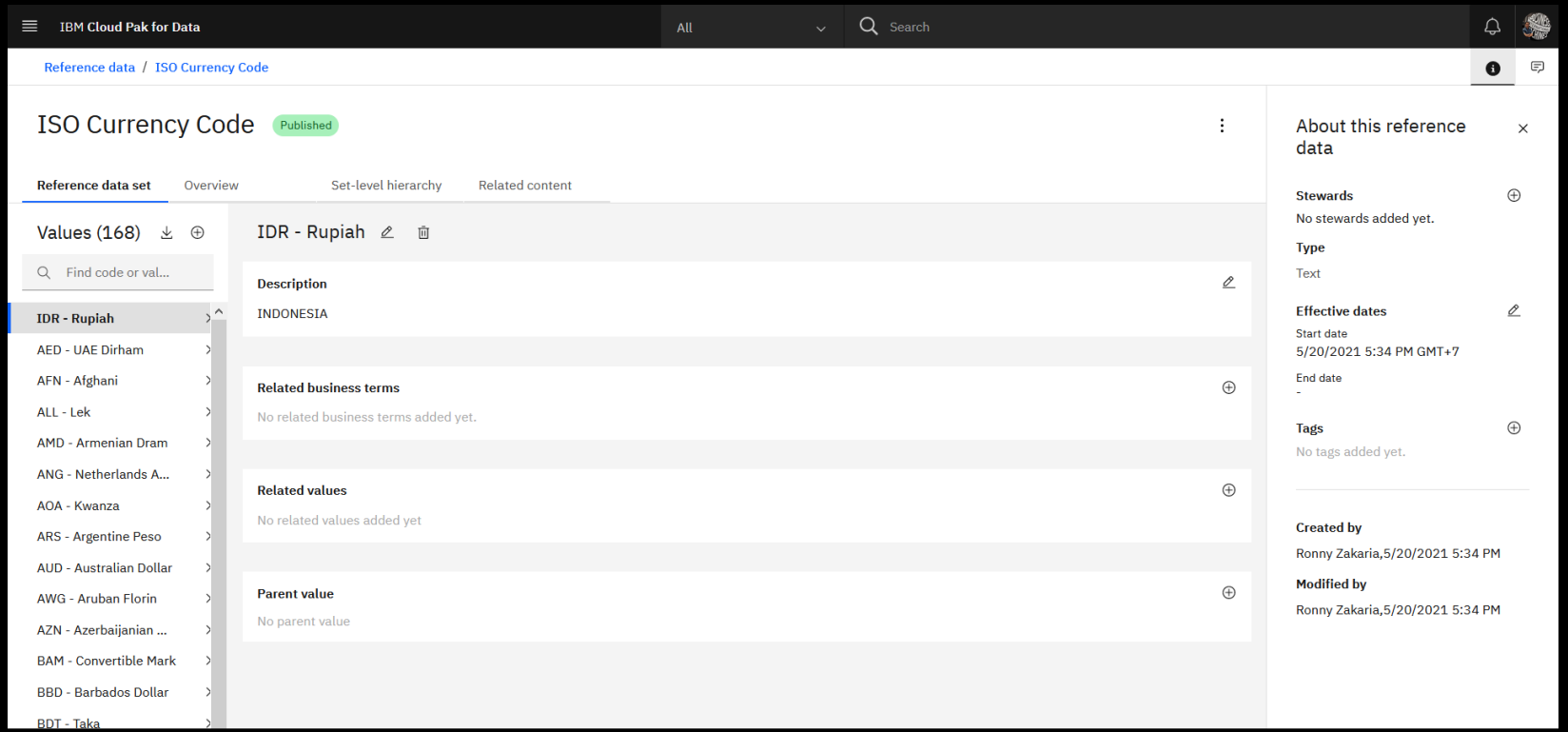


Data Classes represent a set of instructions for detecting and identifying a specific type of data, using pattern or value matching. The Data Discovery services rely upon Data Classes for **application of Rules**.

IBM Knowledge Catalog – Data Classes

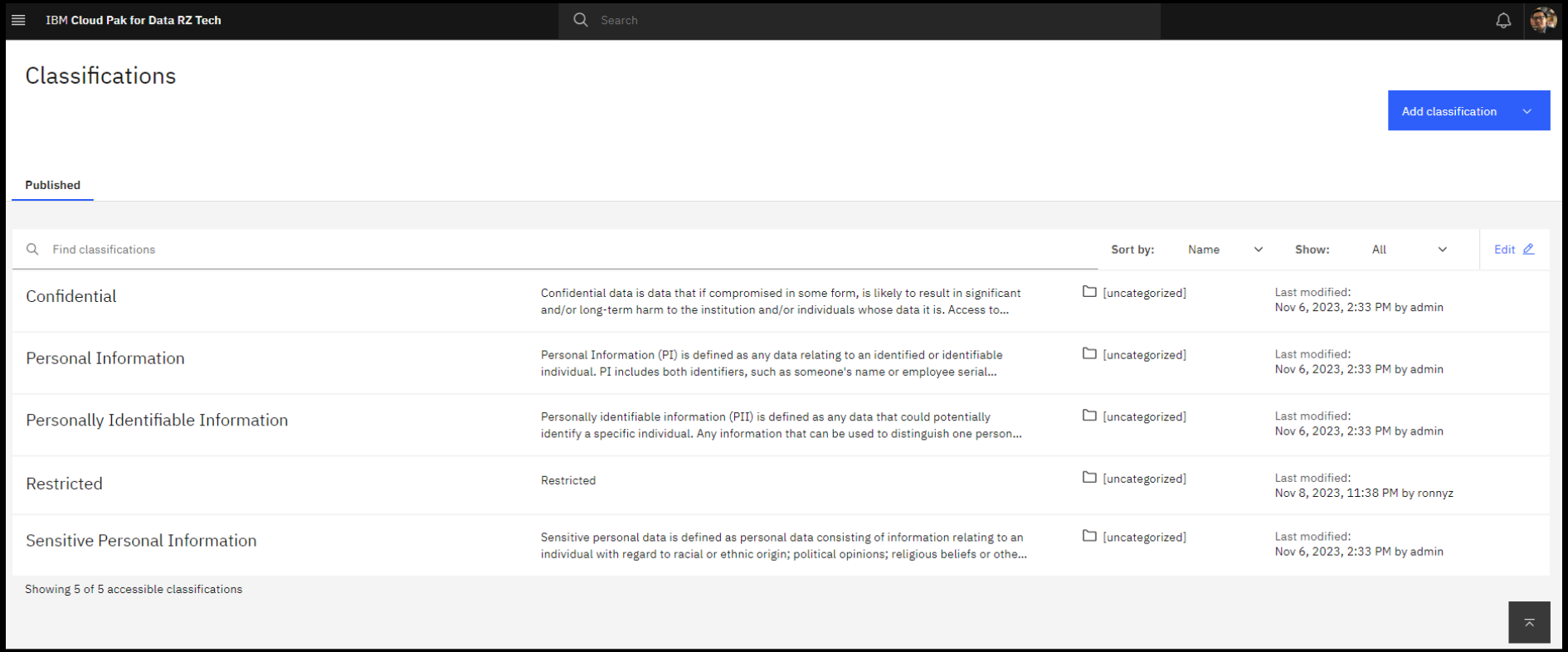


IBM Knowledge Catalog – Reference Data



Reference Data represent a set of allowable or expected values, and are used in conjunction with a Data Class to validate information in the Catalog. Reference Data may be created from the Catalog and values imported from a file.

IBM Knowledge Catalog – Classification



Classifications used to **describe the sensitivity of the data** in data assets. **Each data asset has one classification.** Catalog collaborators assign the classification when they add data assets to a governed catalog. **You can assign classifications to governance artifacts,** such as **business terms, data classes, reference data, policies, and governance rules.**

Agenda

IBM Knowledge Catalog Overview

Delivering Business Ready Data

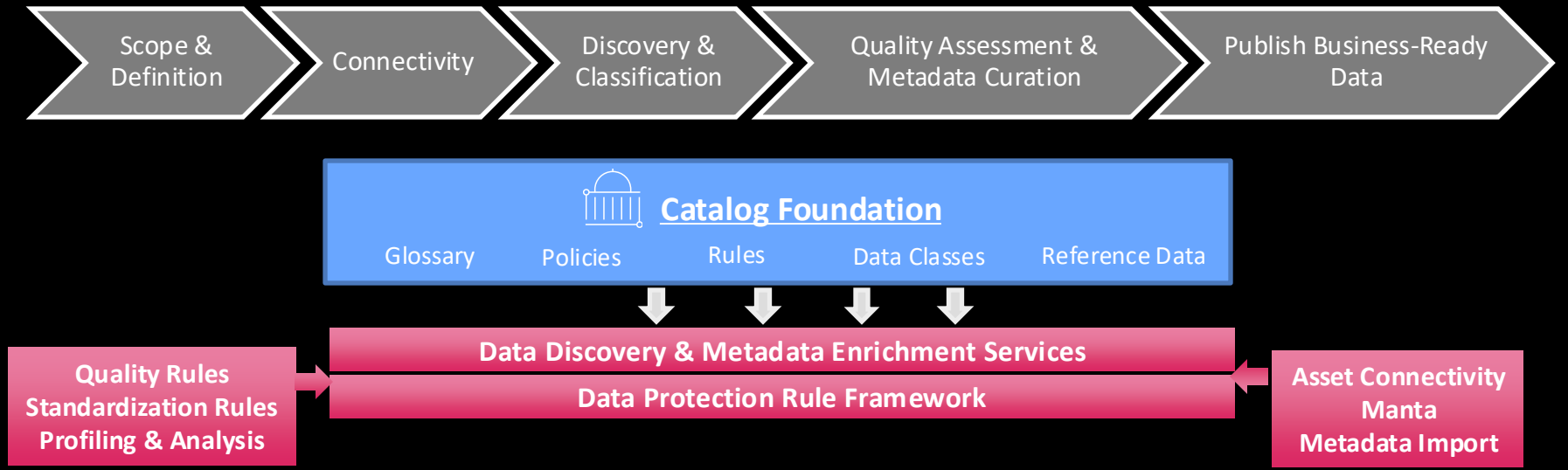
- **Step I – Catalog Foundation.**
- **Step II – Metadata Import and Discovery.**
- **Step III – Delivery of Business-Ready Data.**

Summary & Discussion

IBM Knowledge Catalog

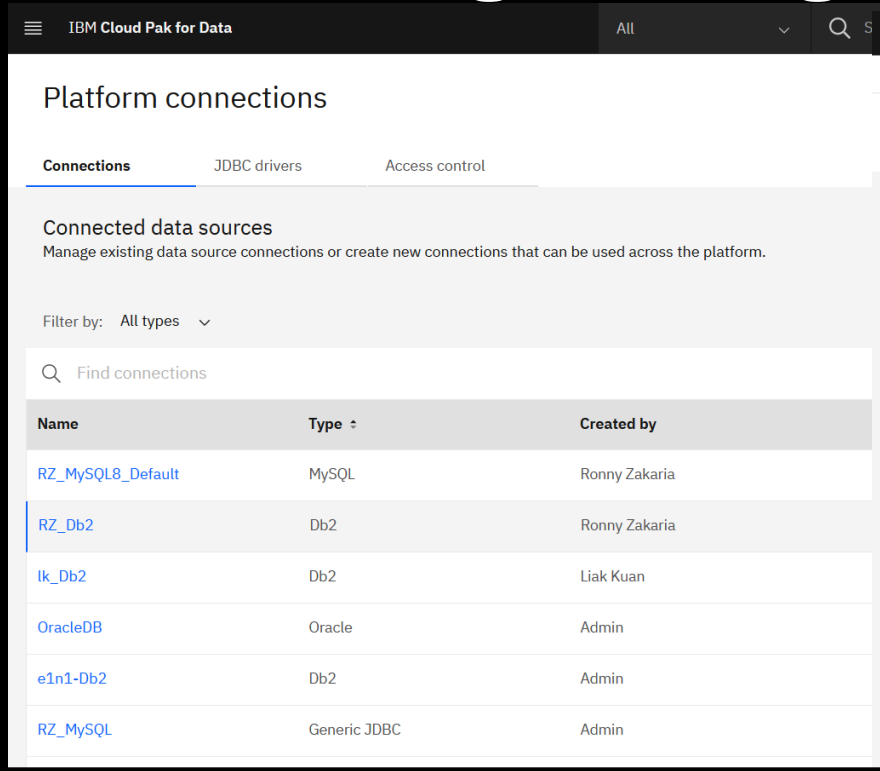
Delivering Business Ready Data to the Enterprise

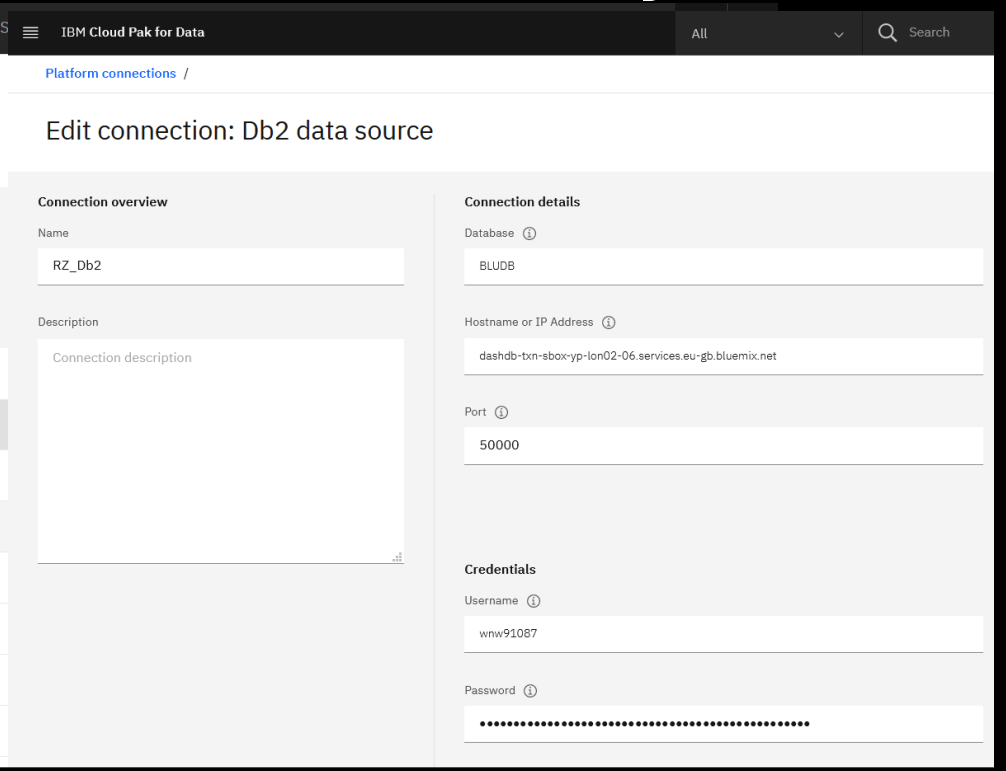
Step II – Metadata Import & Discovery



The second step is to load and represent data in the Catalog, data that is representative and data that is deemed useful and required. During the process, data is scanned and assigned meaning, identification and responsibility.

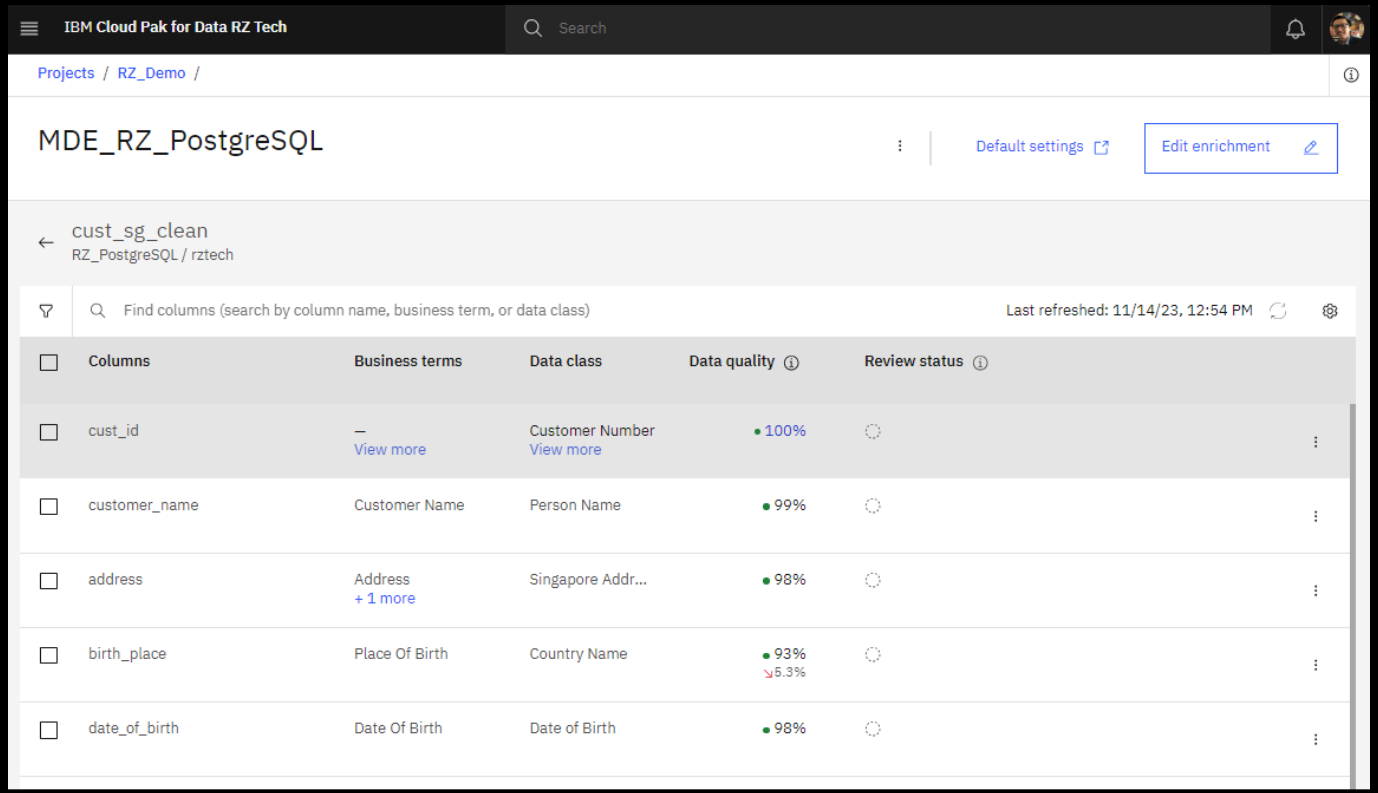
IBM Knowledge Catalog – Data Connectivity





Platform Connections provide a set of parameters and instruction for loading Asset metadata into the Catalog, information about the structure and definition of the Asset. A sub-set of Connections are supported for Discovery.

IBM Knowledge Catalog – Metadata Enrichment



Metadata Enrichment for a specific Database Table, includes the suggested Data Class and Term for each Column, and Data Quality results.

Metadata Enrichment with LLMs

What did we introduce?

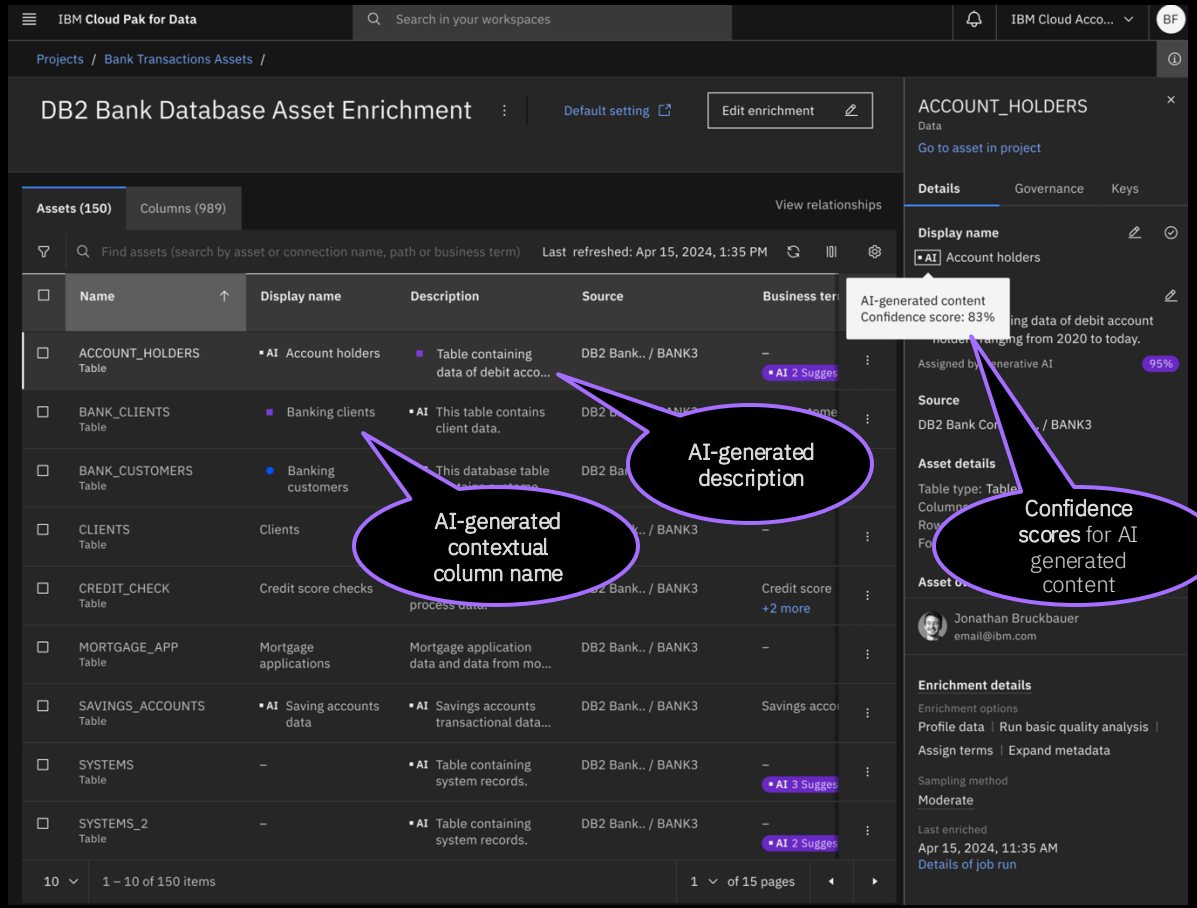
New - Automatically generate meaningful column names and descriptions with context

Enhanced - Assign terms based on semantic meaning and context

Early results - Internal testing showed double the number of correct column mappings (reduction in false positives), and improved precision assignments with LLMs

<u>The value:</u>

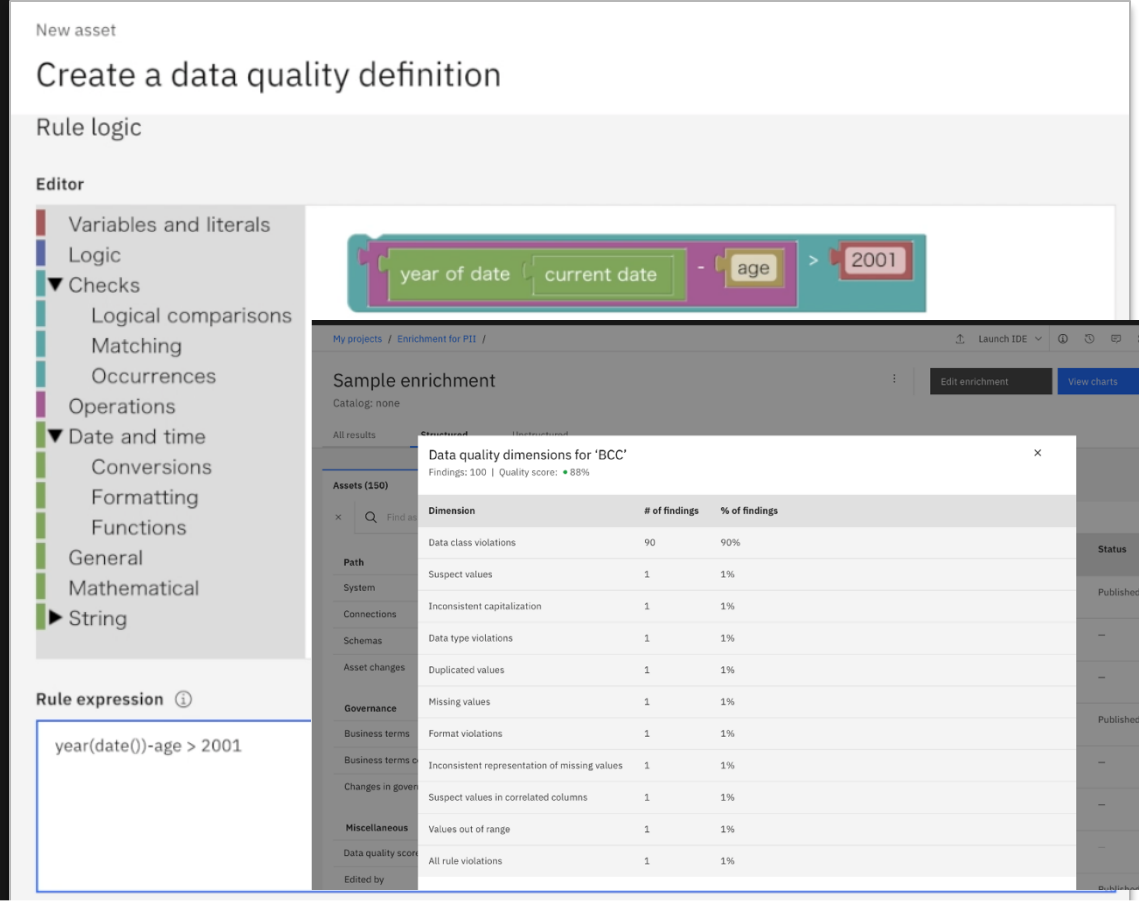
Accelerate data curation through increased accuracy and precision of auto term assignments using AI and LLMs.



Define DQ rule logic & Assign DQ Dimensions, Bind rule to columns

Flexible rule syntax supports rules created by users on legacy product stack

Automatic validation of rule syntax with user tips accelerates new rule development

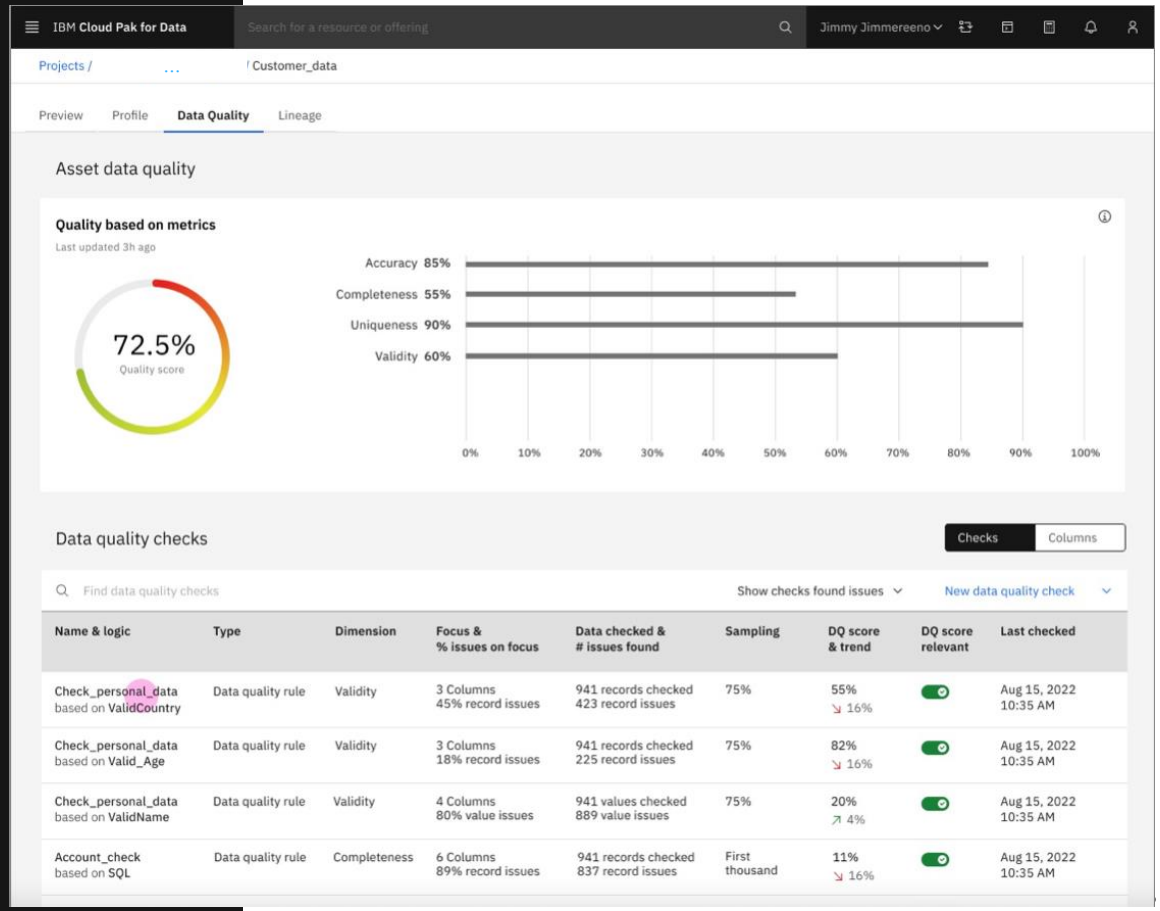


Data Quality Dimensions : there are 10 Default Quality dimension
You can also add your own custom Data Quality rules to be included in the quality score

Built-in DQ Dashboard

Checks view:

- ✓ DQ checks assigned to DQ dimensions
- ✓ De-activate less important DQ checks from DQ score

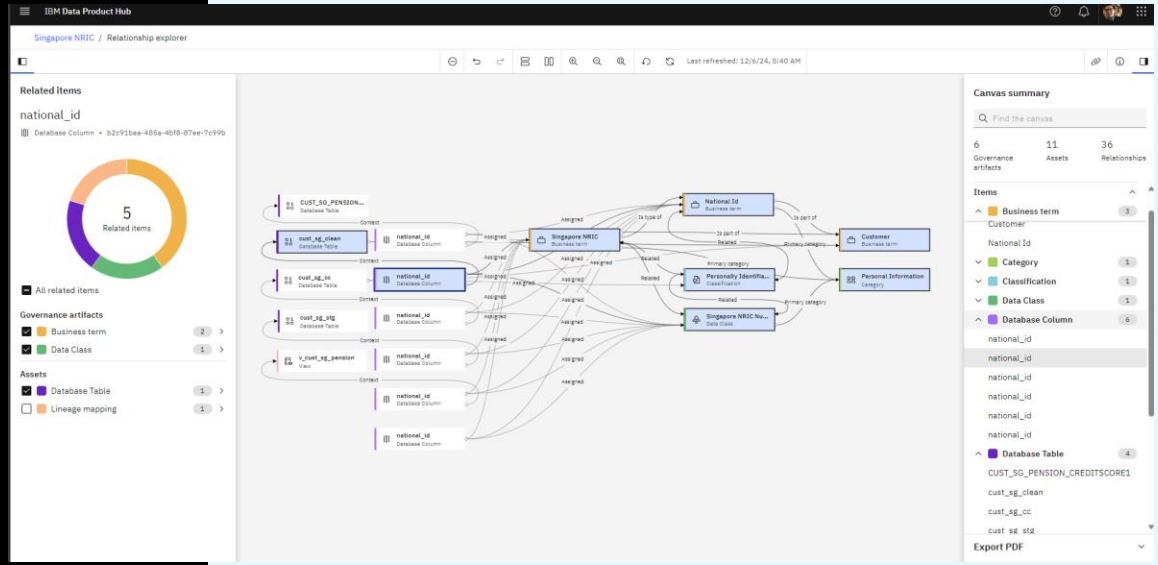


Manage governance artifact relationship

Explore your data landscape and relationships through visualization

Relationship explorer is a visualization tool of the knowledge graph relationships. It enables users to graphically navigate between business and technical assets through their relationships.

- **Data consumers** can find relevant data assets and their relationships with business concepts and other data assets. They can understand the impact and importance of data governance.
- **Data stewards** can explore asset relationships for impact analysis, audit metadata and investigate data governance.



"When you're crafting a business term and putting it together, at the end of the day, it's not in isolation. It's there for a reason. And I'm seeing "the reason".

Agenda

IBM Knowledge Catalog Overview

Delivering Business Ready Data

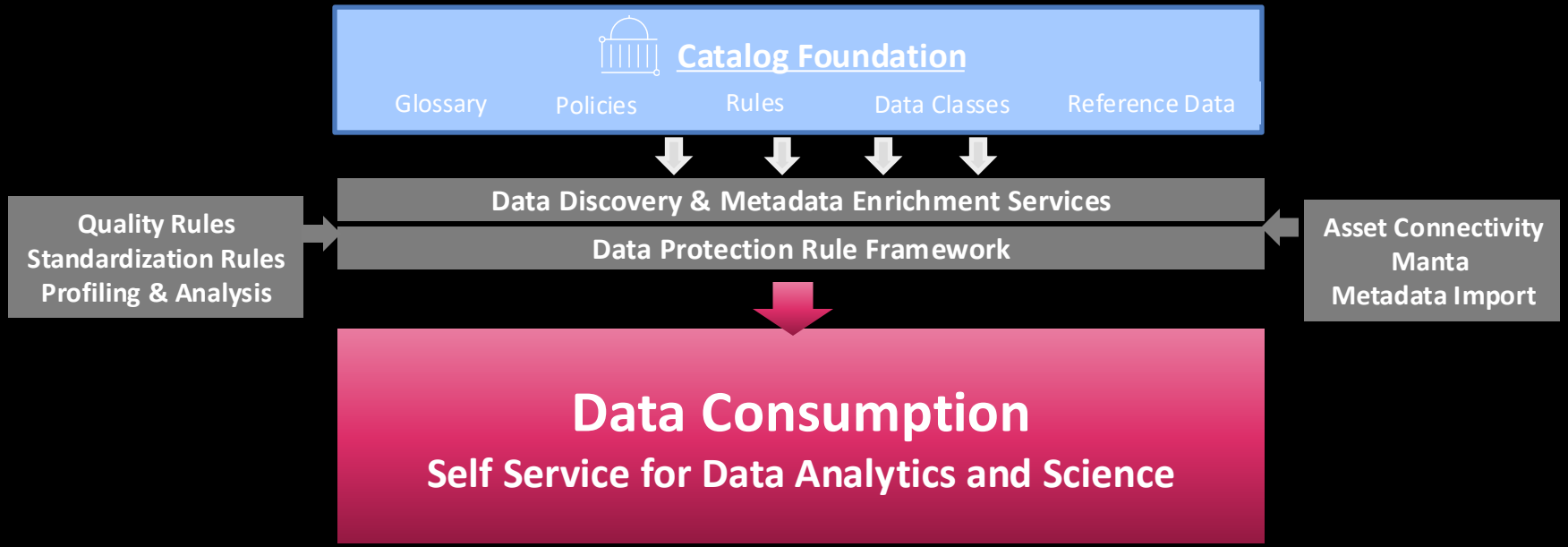
- **Step I – Catalog Foundation.**
- **Step II – Metadata Import and Discovery.**
- **Step III – Delivery of Business-Ready Data.**

Summary & Discussion

IBM Knowledge Catalog

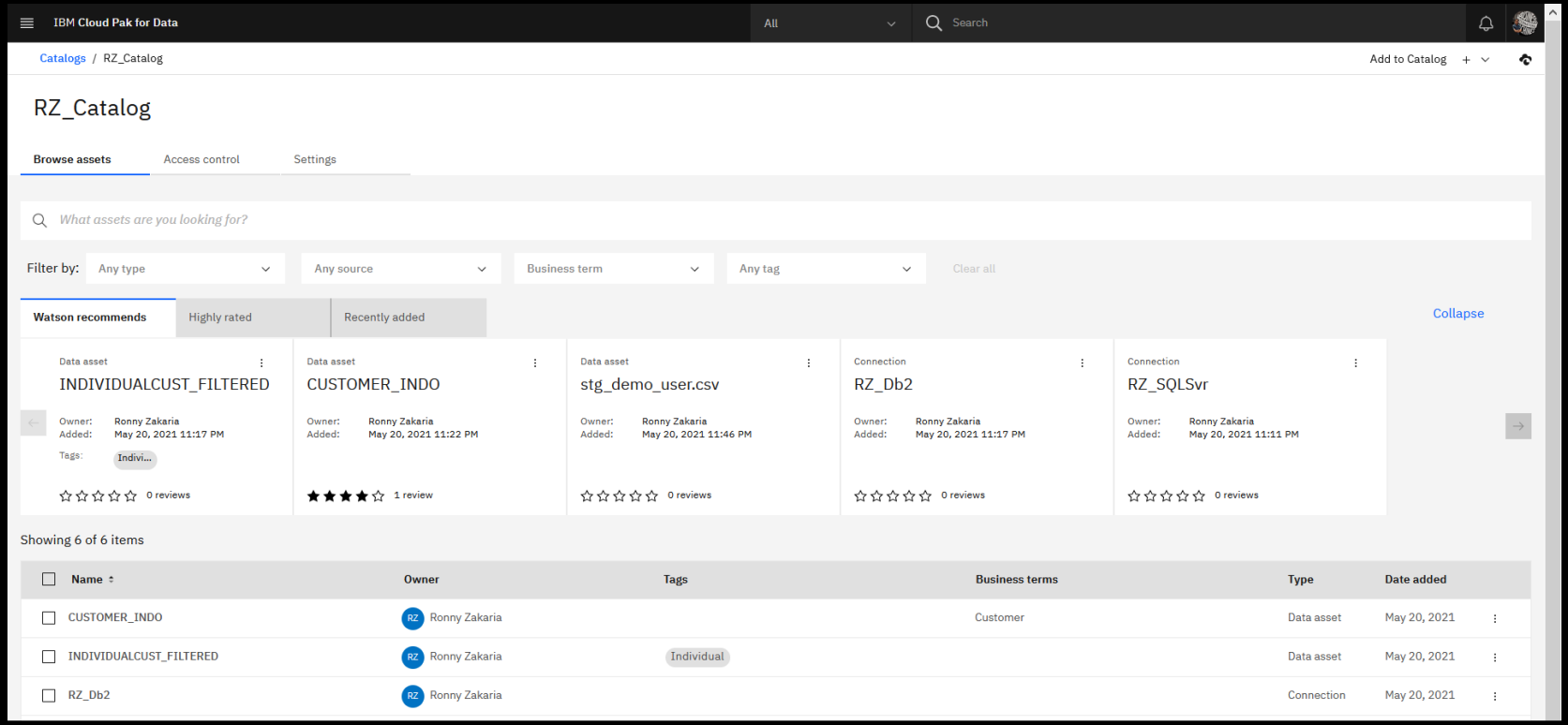
Delivering Business Ready Data to the Enterprise

Step III – Delivery of Business Ready Data



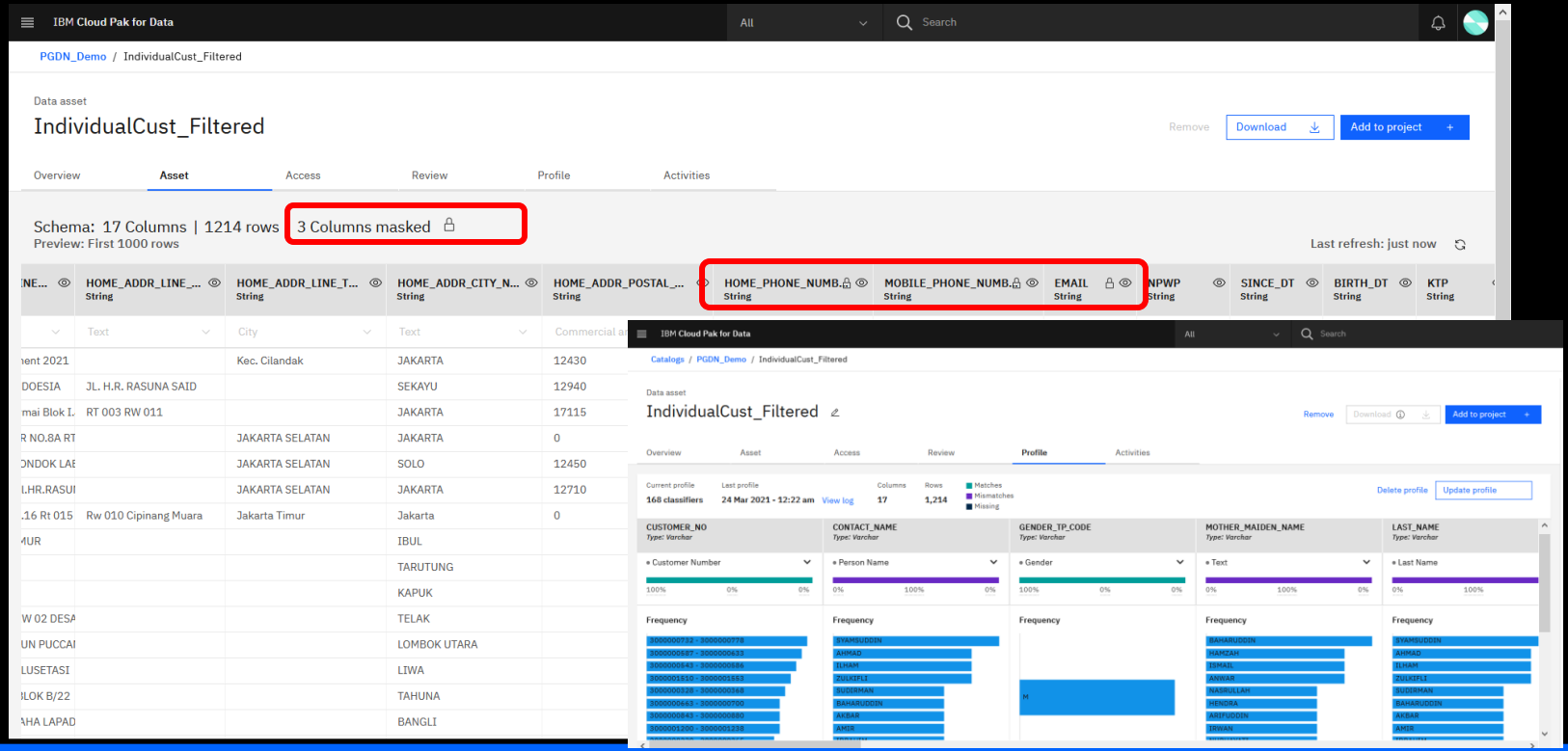
The Final step is to **allow users to search and explore business-ready data, gaining understanding, building trust and ensuring responsibility to data.** Understanding how to use data to drive new Analytics or Data Science.

Catalog



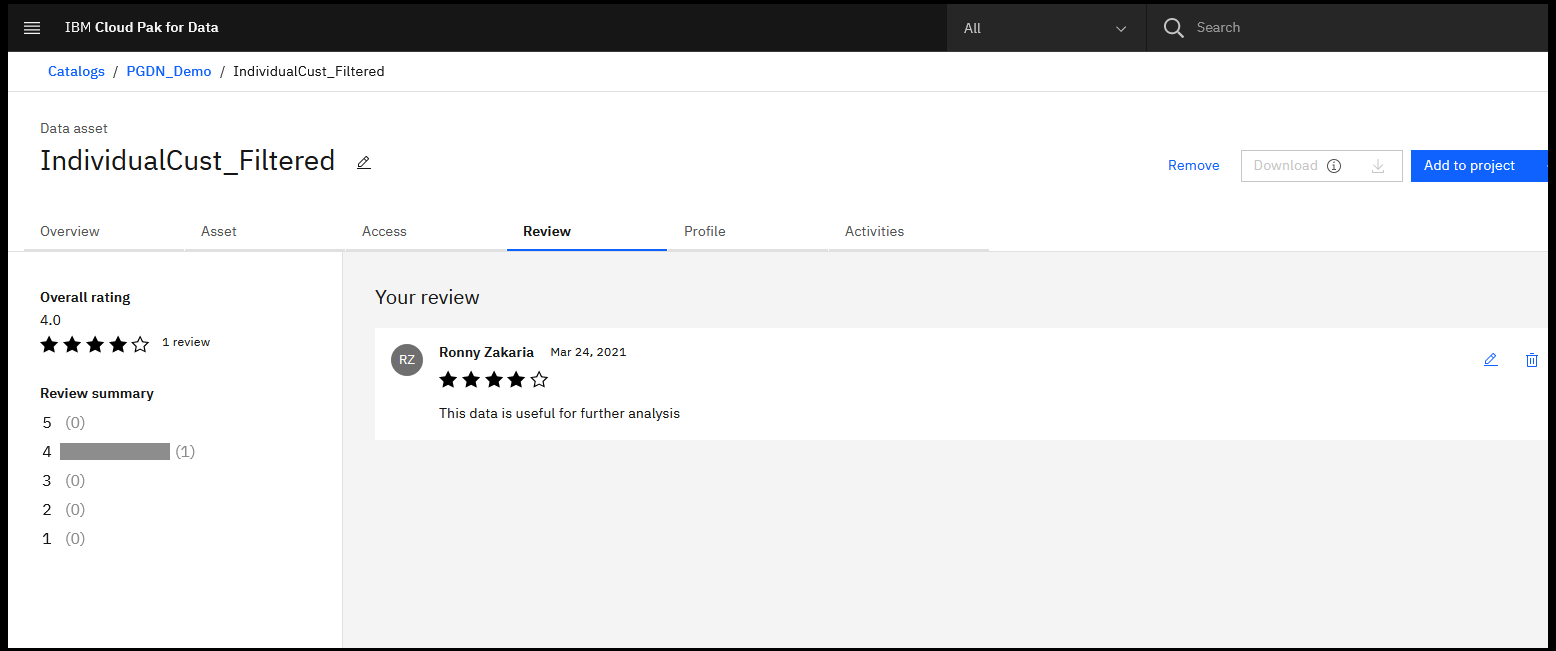
Catalog connects data and knowledge with the people who need to use it.
A **catalog** is where you share assets across your enterprise. A catalog includes **people** and the **assets** that they need to find

IBM Knowledge Catalog – Data Preview and Analysis



Preview a Data Set, understanding the referential set of data and its quality scoring or frequency distribution. Load data into a Project for further refinement or visualization of the data.

IBM Knowledge Catalog – Comment and Ratings



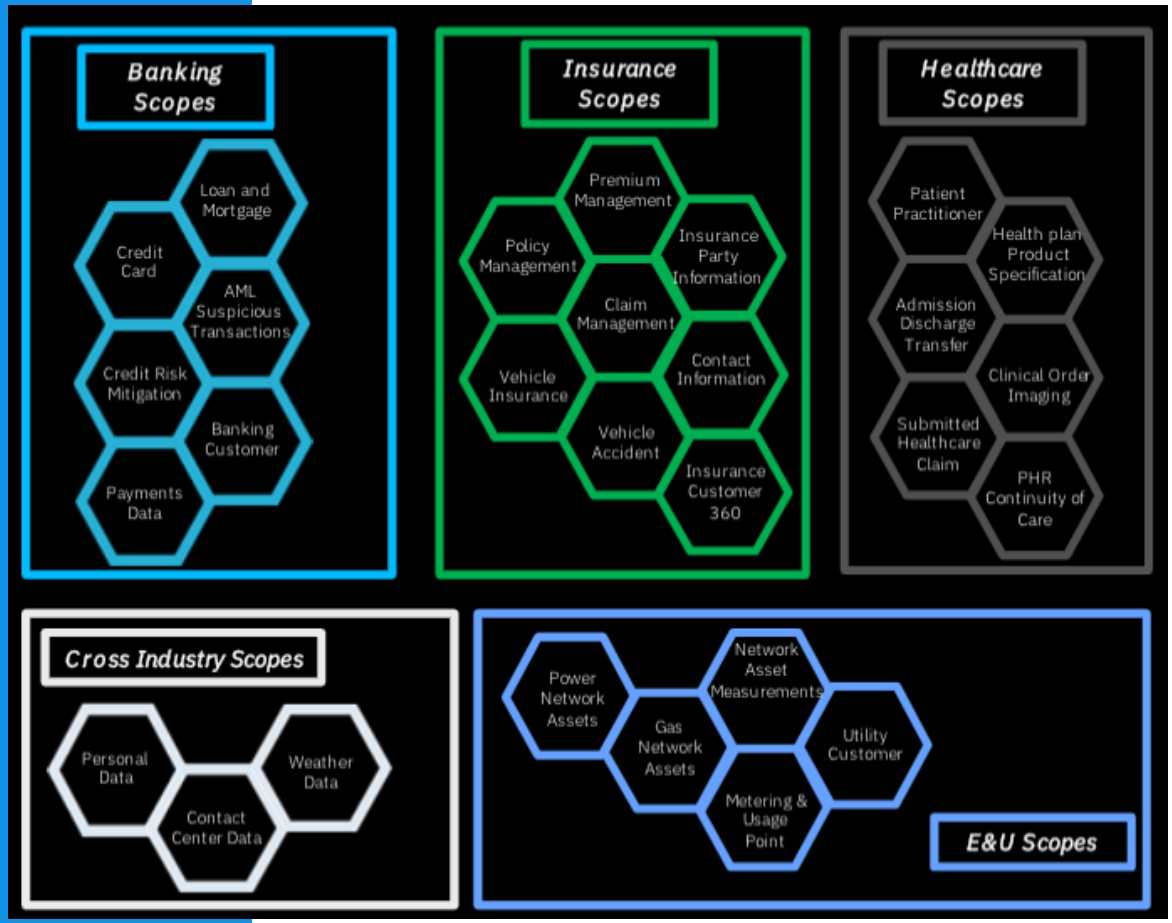
Collaborate amongst users by leaving a comment or assignment of a Rating to an Asset.

Knowledge Accelerator

Accelerate regulatory compliance

Align concepts from industry regulations and standards with your business data

- Set of business scopes for each supported industry that are smaller and more manageable in size
- New cross industry scopes
- Easy to install
- Available at no extra cost with CPD software and SaaS Enterprise plan



Agenda

IBM Knowledge Catalog Overview

Delivering Business Ready Data

- **Step I – Catalog Foundation.**
- **Step II – Metadata Import and Discovery.**
- **Step III – Delivery of Business-Ready Data.**

Summary & Discussion

Summary

- IBM Knowledge Catalog provides **Enterprise Data Governance**, **Data Quality** and **Data Consumption**

- IBM Knowledge Catalog **serves multiple personas**:
  Data Citizen, Information Analyst, Data Steward, Quality Analyst

- IBM Knowledge Catalog + Manta provide **broad connector for Data Lineage**

- IBM Knowledge Catalog + Manta **built on Modernized Architecture** that is **Ready to run on Any Cloud** including **Private Cloud on Premise**

© 2019 IBM Corporation