Informatica

26 Aug, 2025

Enhancing Enterprise GenAI with MDM and Data Governance

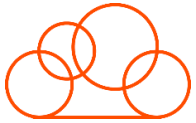
- David Hendrawirawan, Principal Consultant, IPS
- Troy Coleman, Solution Architect, IPS
- Alagar Pitchai, Solution Architect, IPS

Where data & AI come to LIFE

Housekeeping Tips

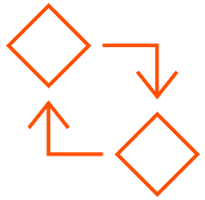
- Today's Webinar is scheduled for 1 hour
- The session will include a webcast and then your questions will be answered live at the end of the presentation
- All dial-in participants will be muted to enable the speakers to present without interruption
- Questions can be submitted to "All Panelists" via 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

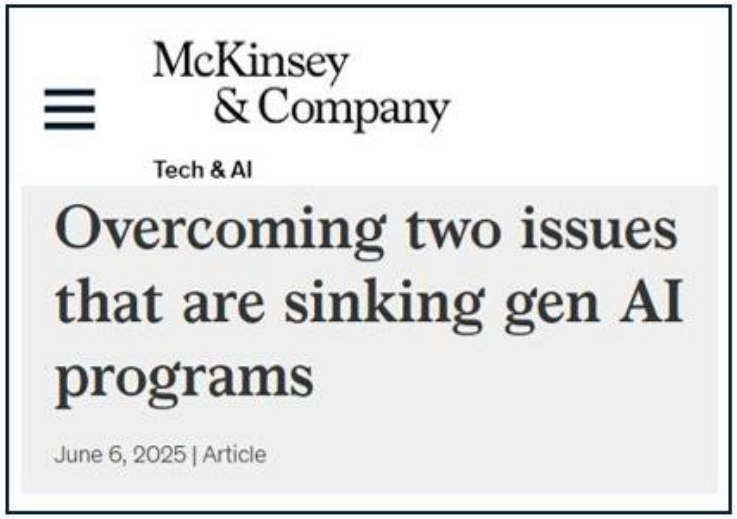
The information being provided today is for informational purposes only. The development, release, and timing of any Informatica product or functionality described today remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision.

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.

Agenda

1. Generative AI Challenges for Enterprise Scale
2. Enabling Patterns for Enhancing Enterprise GenAI
3. Overview of GenAI Enablers in CDGC
4. Overview of GenAI Enablers in MDM
5. Case Study

The State of Generative AI for Enterprise

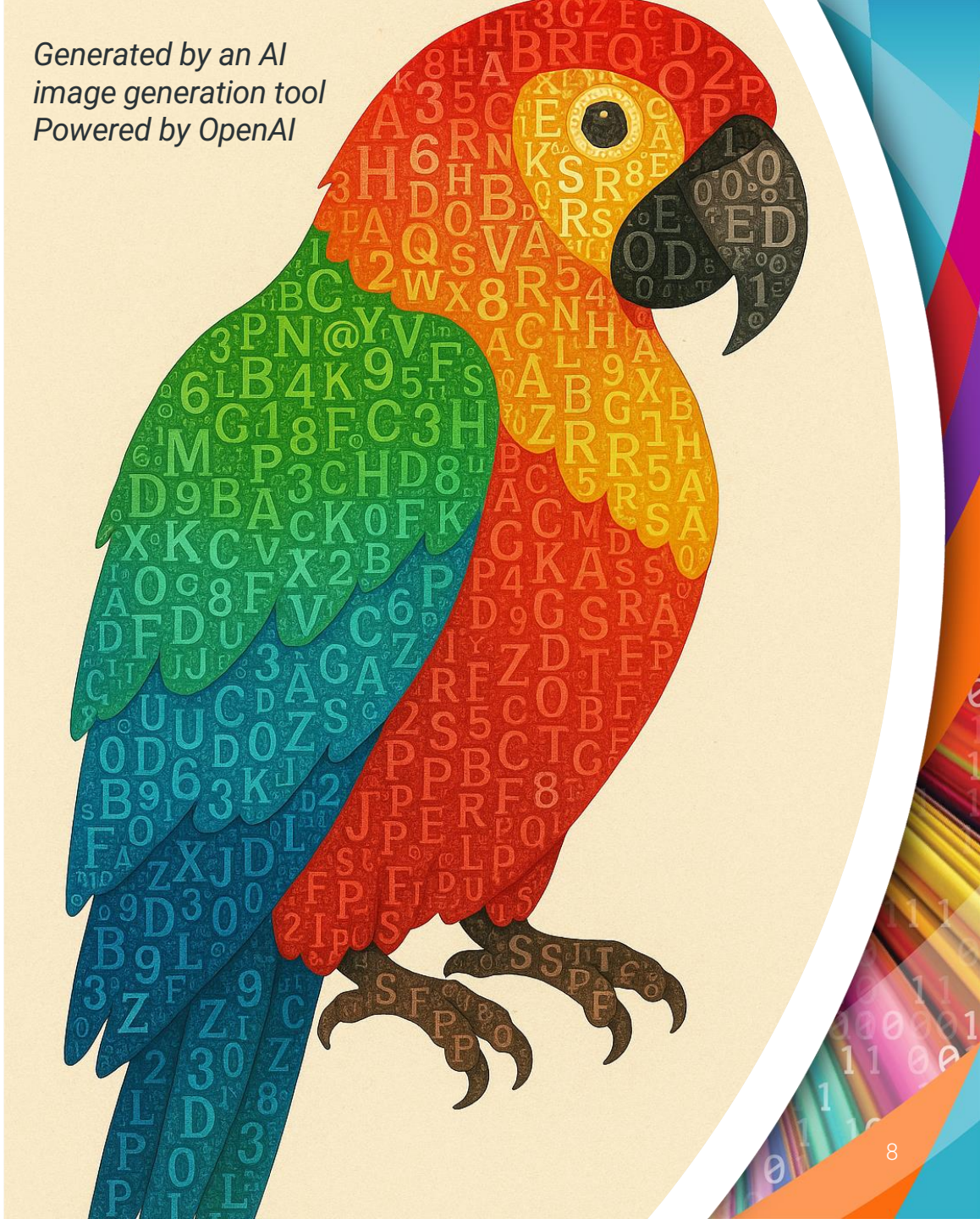


The “Stochastic Parrot”

The term "stochastic parrot" refers to a critique of large AI language models, highlighting that they generate text by statistically predicting the next word based on patterns learned from its training data, without true understanding.

These models essentially "parrot" or mimic human language probabilistically rather than reasoning or comprehending meaning.

This concept raises important questions about the ethical use, transparency, and limitations of AI-generated content.

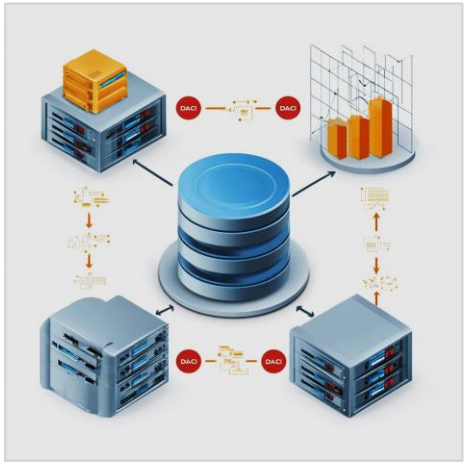


Generated by an AI image generation tool Powered by OpenAI

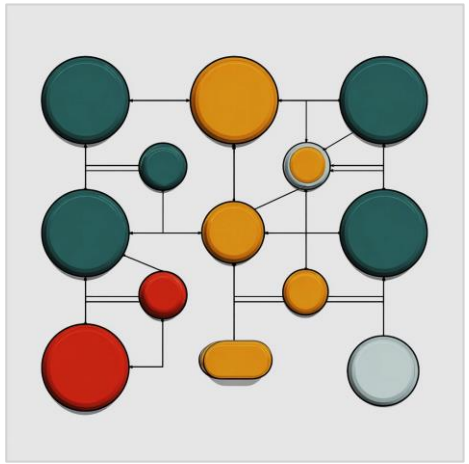
The Three Sources of Enterprise Knowledge & Intelligence

Unstructured Data

Golden Record



Metadata



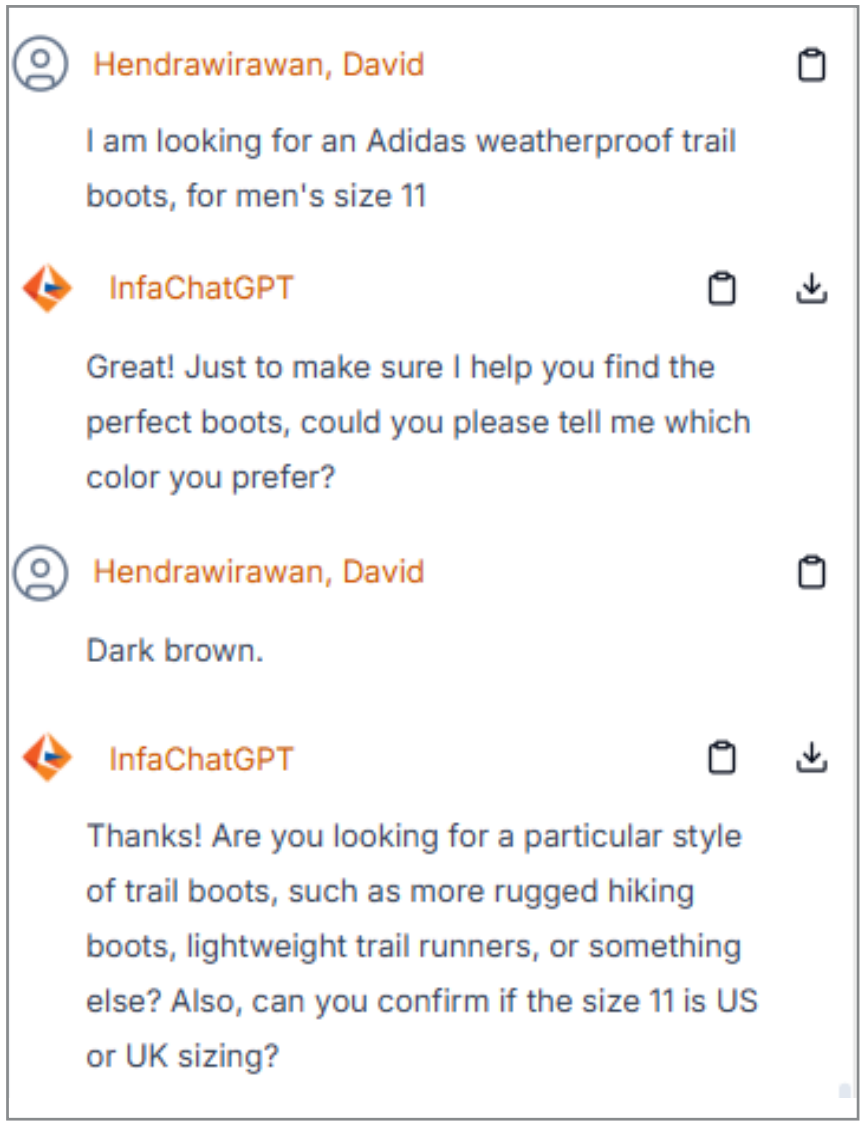
Fluid - Adaptability

Solid - Authoritative

Frame – Structural Integrity

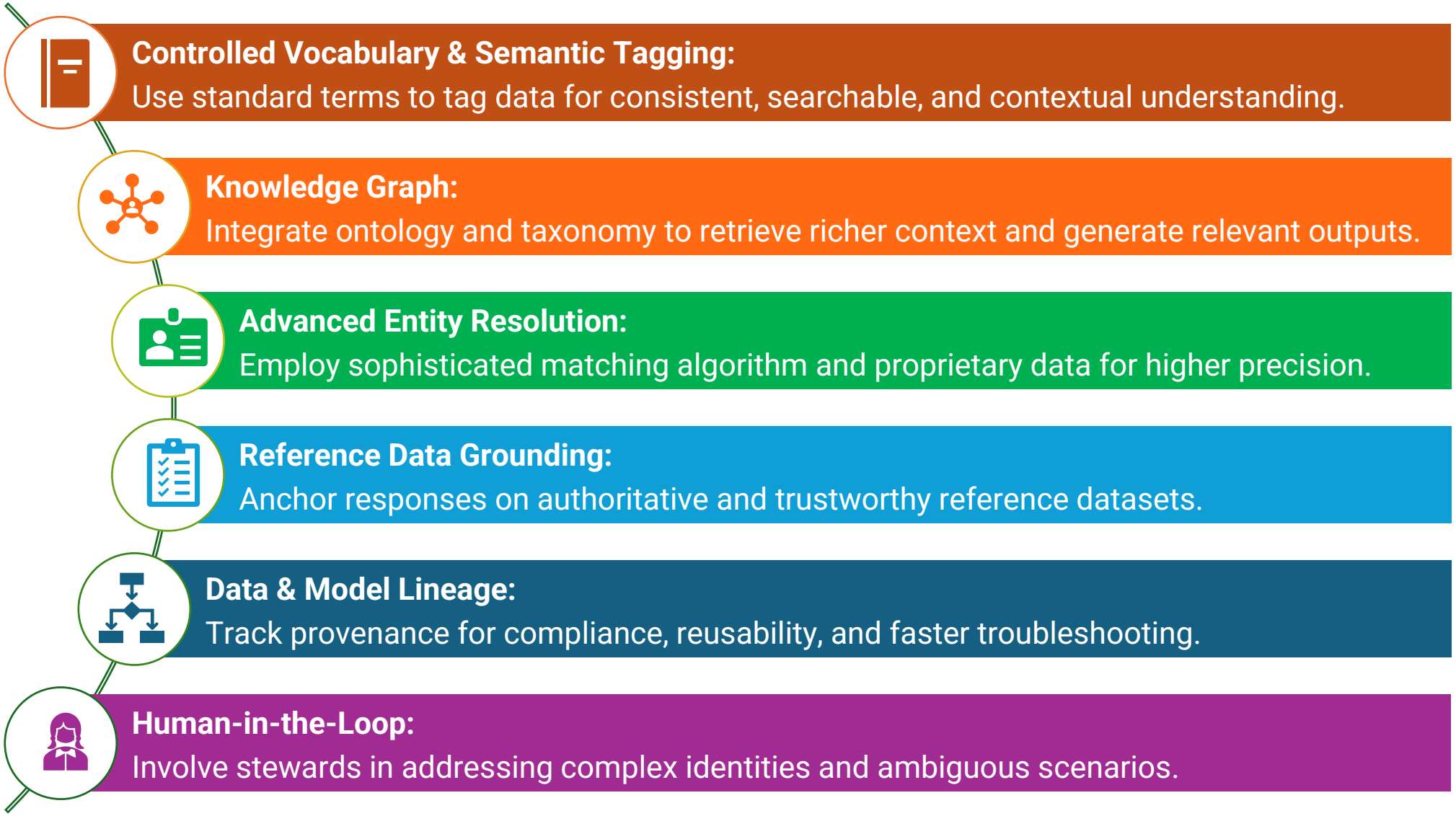
Informatica

Case Study: Sports Apparel Recommendation Chatbot

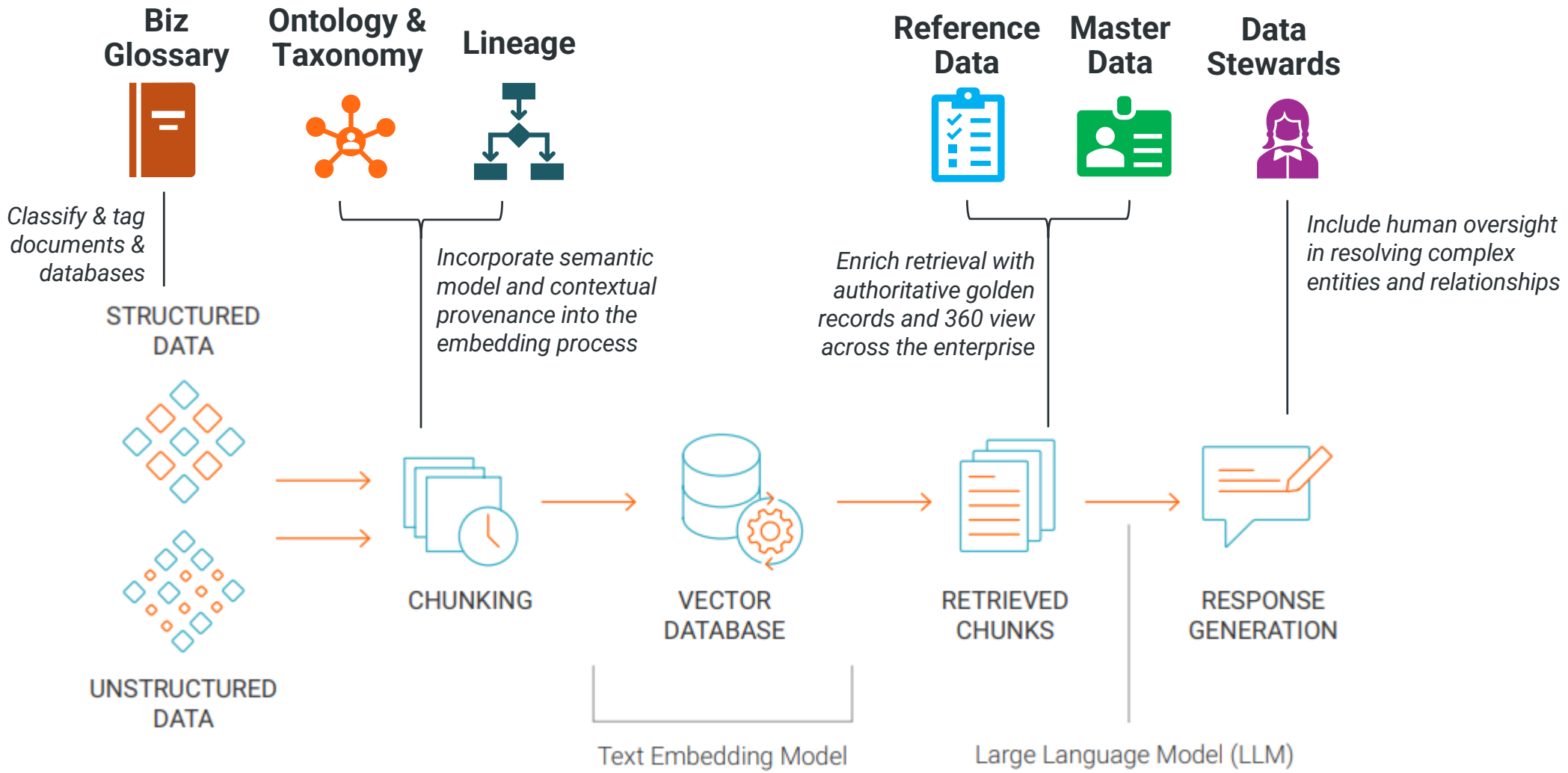


Informatica

Enabling Patterns to Enhance Generative AI



Example: CDGC and MDM-Enhanced RAG



Informatica

How CDGC Enhances GenAI

Where data & AI come to LIFE

CDGC Enhances GenAI…

- QUALITY – Improving AI Outputs with Meaning, Relationship, and Structure
  - Business Terms – AI-Mapped to Technical Metadata
  - Taxonomy – The Hierarchy of Domains, Sub-Domains, and Business Terms
  - Ontology – The Relationship among Business Terms
- RESPONSIBILITY – Enforcing AI Controls with Lineage Discovery and Policy Automation
  - PII Discovery and Auto-Classify Highly Sensitive Data
  - Apply Data Quality Rules Consistently
  - Govern the Approval Process for AI Data Usage
  - Enforce Controls Through an Enterprise Policy Metadata Store
- AGILITY – Speeding AI Projects from Idea to Production with Transparency
  - Publish Certified, AI-Ready Data Assets (Including Data Marketplace Integration)
  - Clarify Governance Accountability Through Data Stewardship Assignment
  - Accelerate with AI Model Lineage and Data Flow Visibility

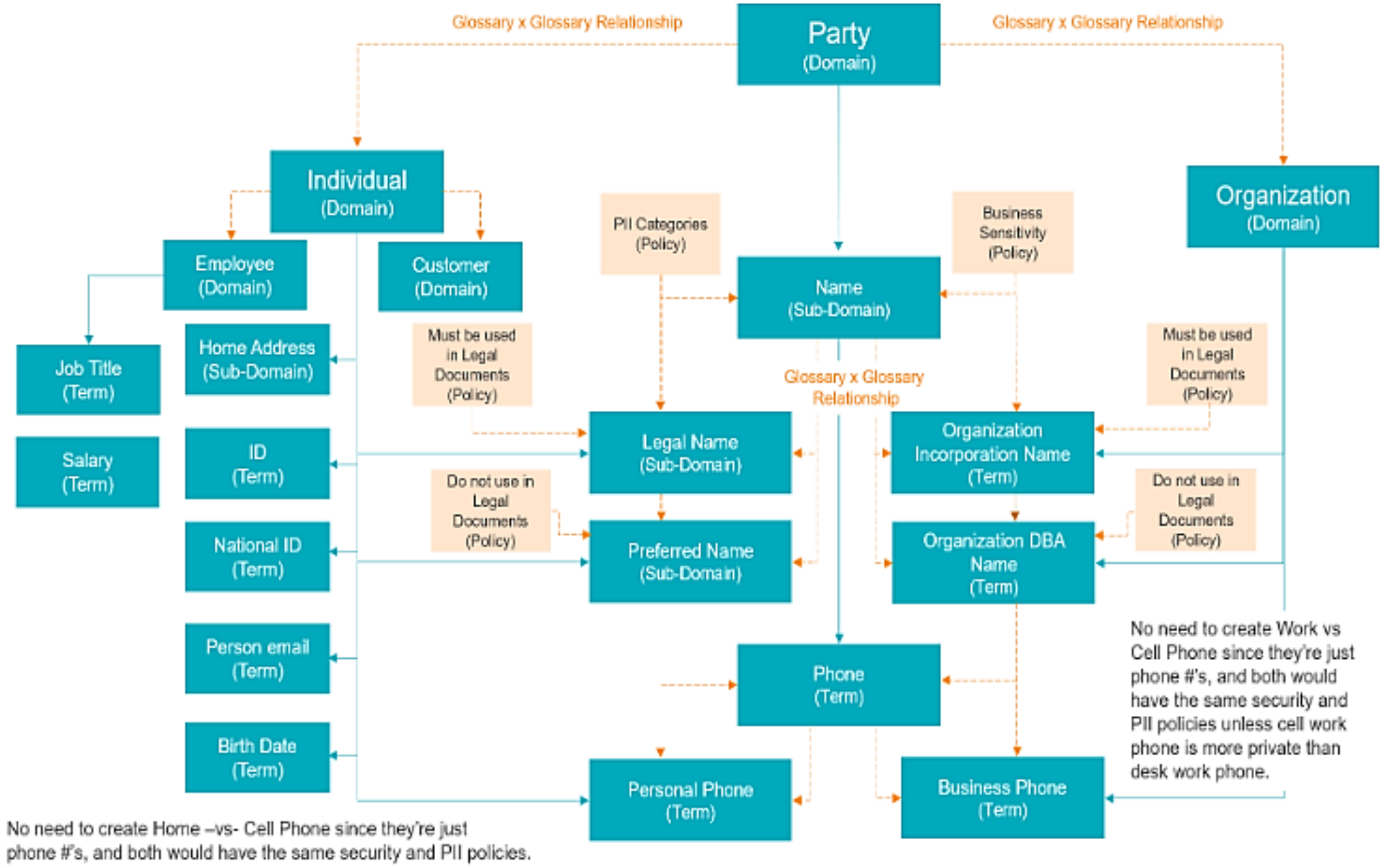
CDGC Enhances GenAI QUALITY

Business Terms – AI-Mapped to Technical Metadata

Taxonomy – The Hierarchy of Domains, Sub-Domains, and Business Terms

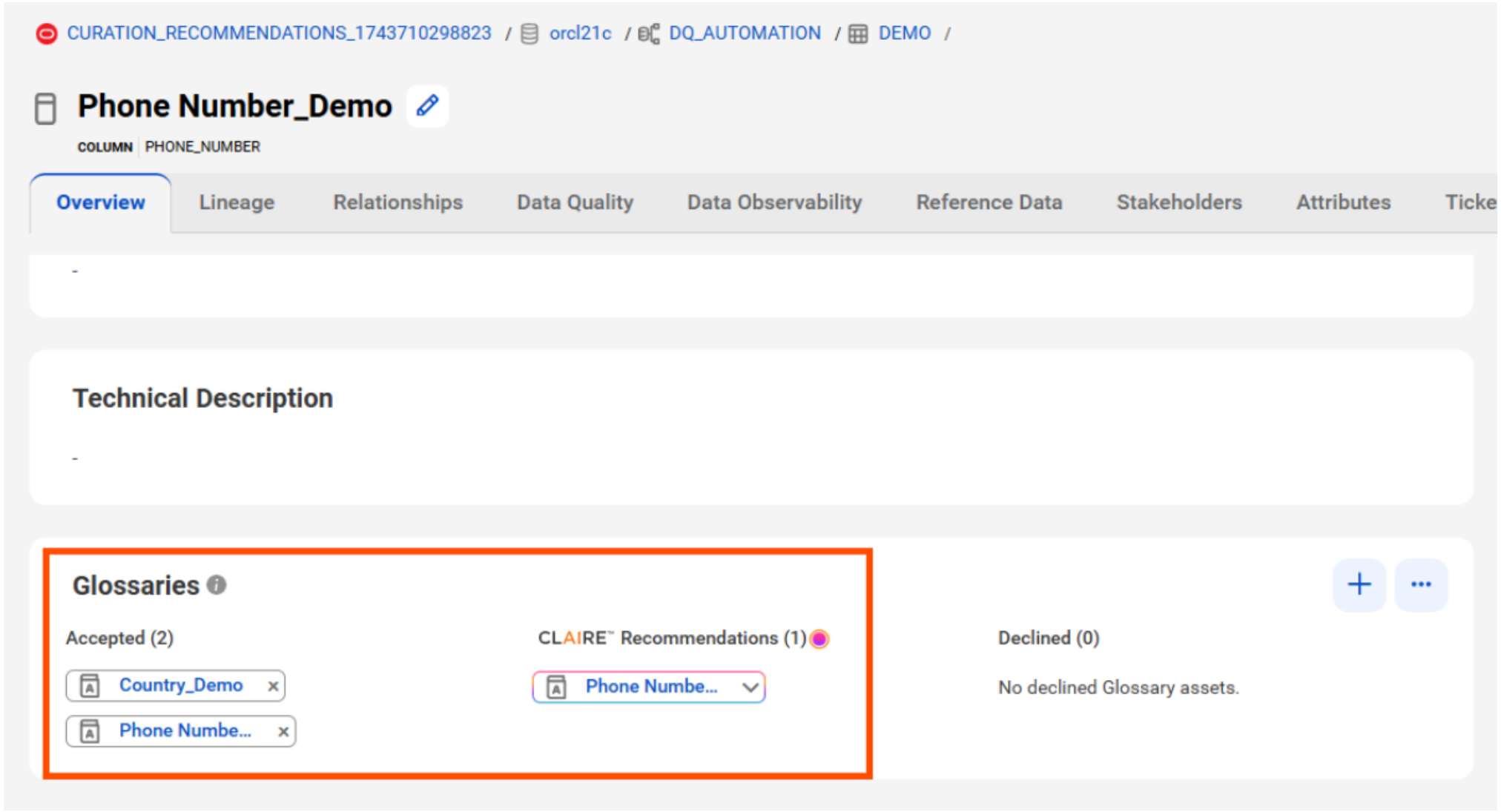
Ontology – The Relationship among Business Terms

Example Biz Terms for Customer Domain

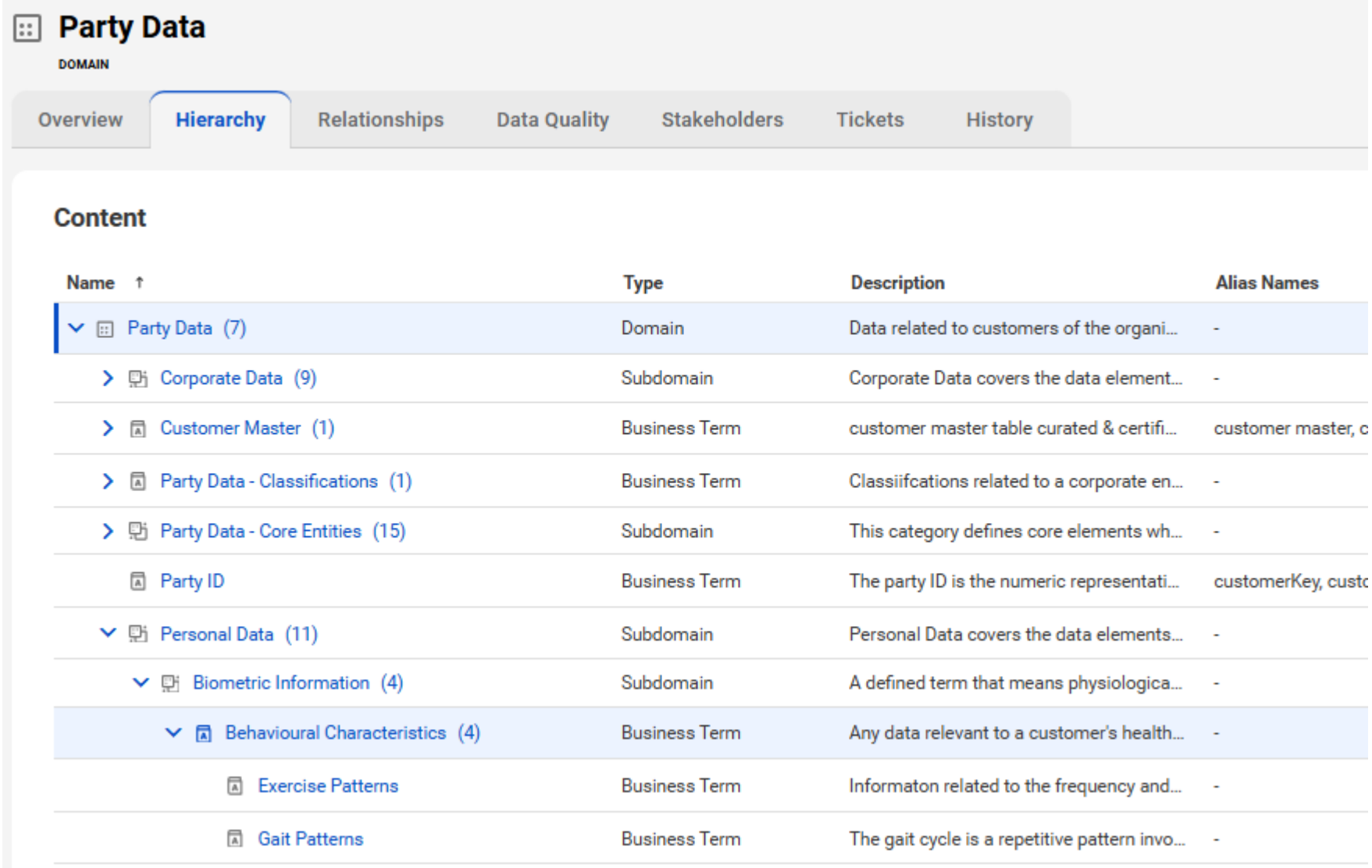


Source: http://velocity.informatica.com/stages-data-governance/dgc-plan/dgc-plan-bestpractices/950-glossary-design-bp

Business Terms – AI-Mapped to Technical Metadata

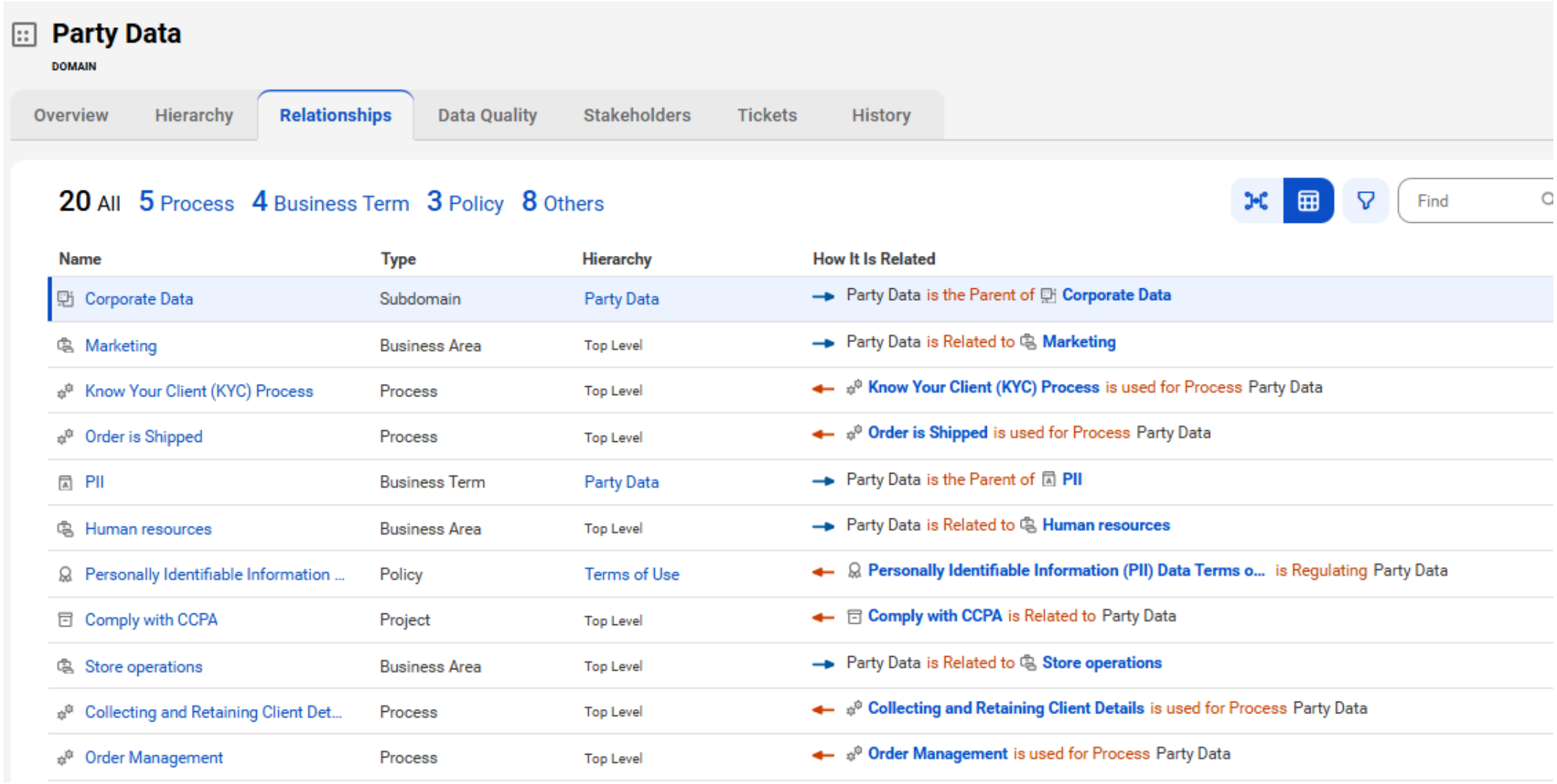


Taxonomy of Business Terms – Customer Domain

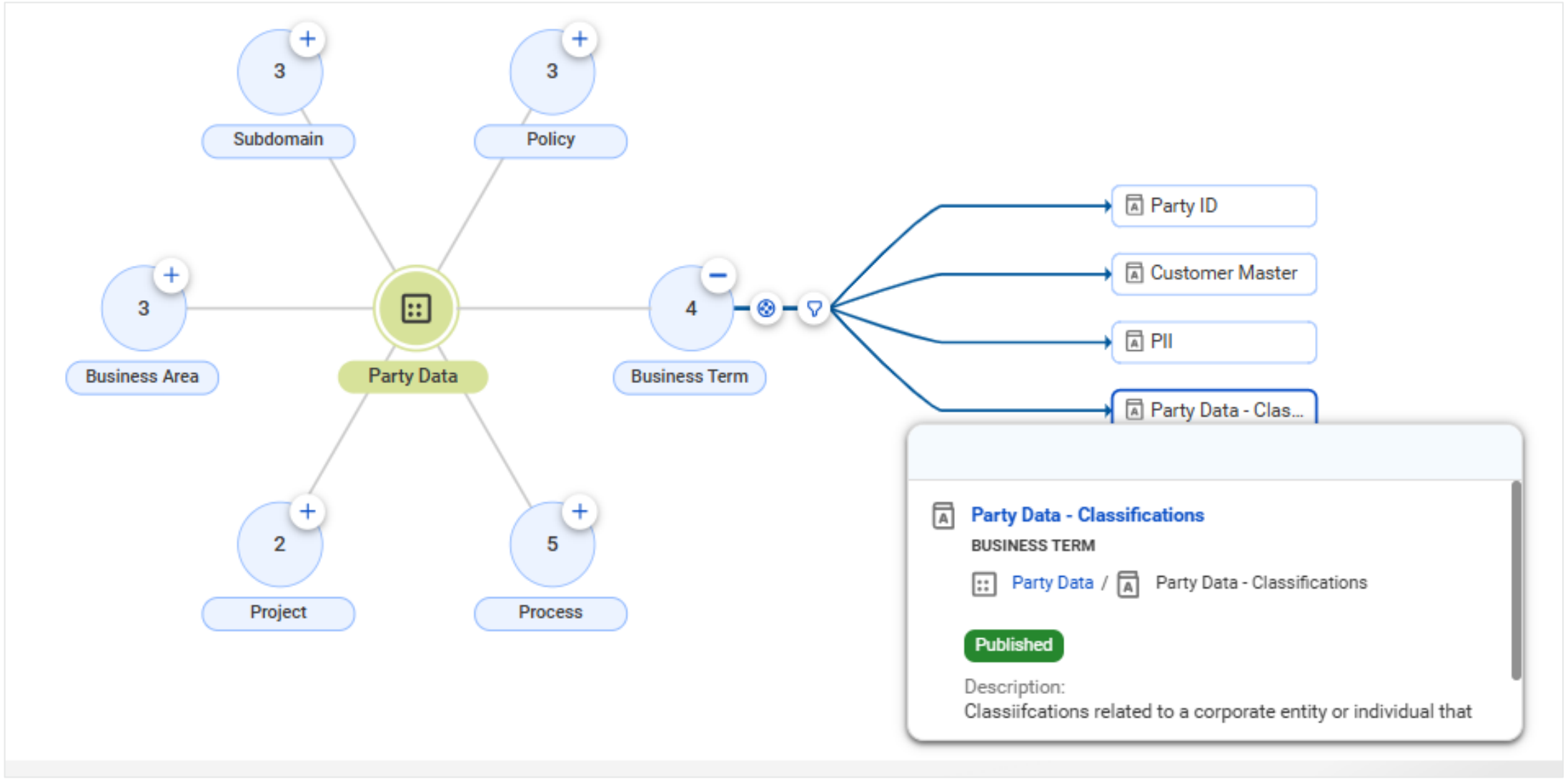


Name	Type	Description	Alias Names
Party Data (7)	Domain	Data related to customers of the organi...	-
Corporate Data (9)	Subdomain	Corporate Data covers the data element...	-
Customer Master (1)	Business Term	customer master table curated & certifi...	customer master, c
Party Data - Classifications (1)	Business Term	Classiifcations related to a corporate en...	-
Party Data - Core Entities (15)	Subdomain	This category defines core elements wh...	-
Party ID	Business Term	The party ID is the numeric representati...	customerKey, custc
Personal Data (11)	Subdomain	Personal Data covers the data elements...	-
Biometric Information (4)	Subdomain	A defined term that means physiologica...	-
Behavioural Characteristics (4)	Business Term	Any data relevant to a customer's health...	-
Exercise Patterns	Business Term	Informaton related to the frequency and...	-
Gait Patterns	Business Term	The gait cycle is a repetitive pattern invo...	-

Ontology of Business Terms



Related Governance Assets – Party Data



CDGC Enhances GenAI RESPONSIBILITY

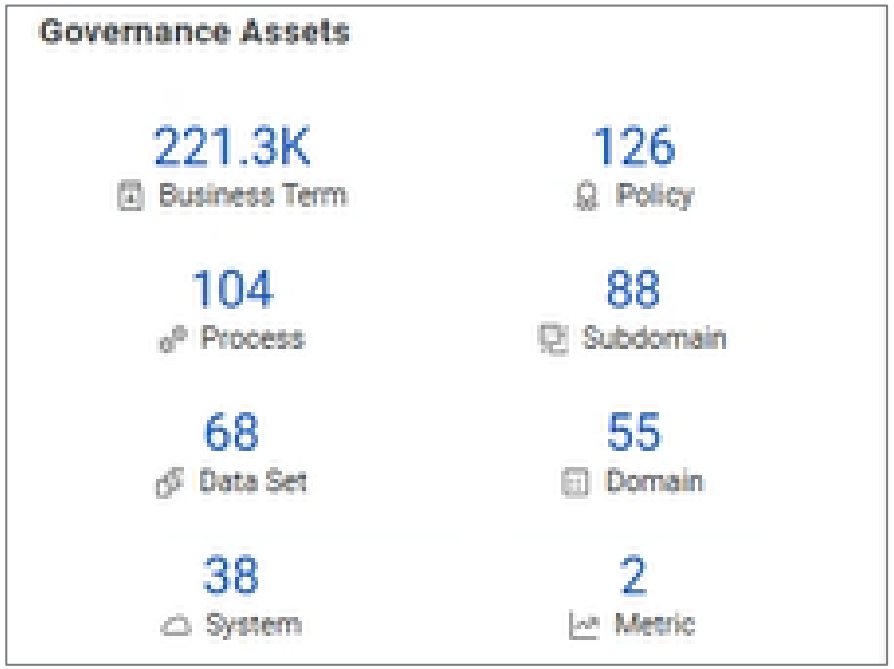
PII Discovery and Auto-Classify Highly Sensitive Data

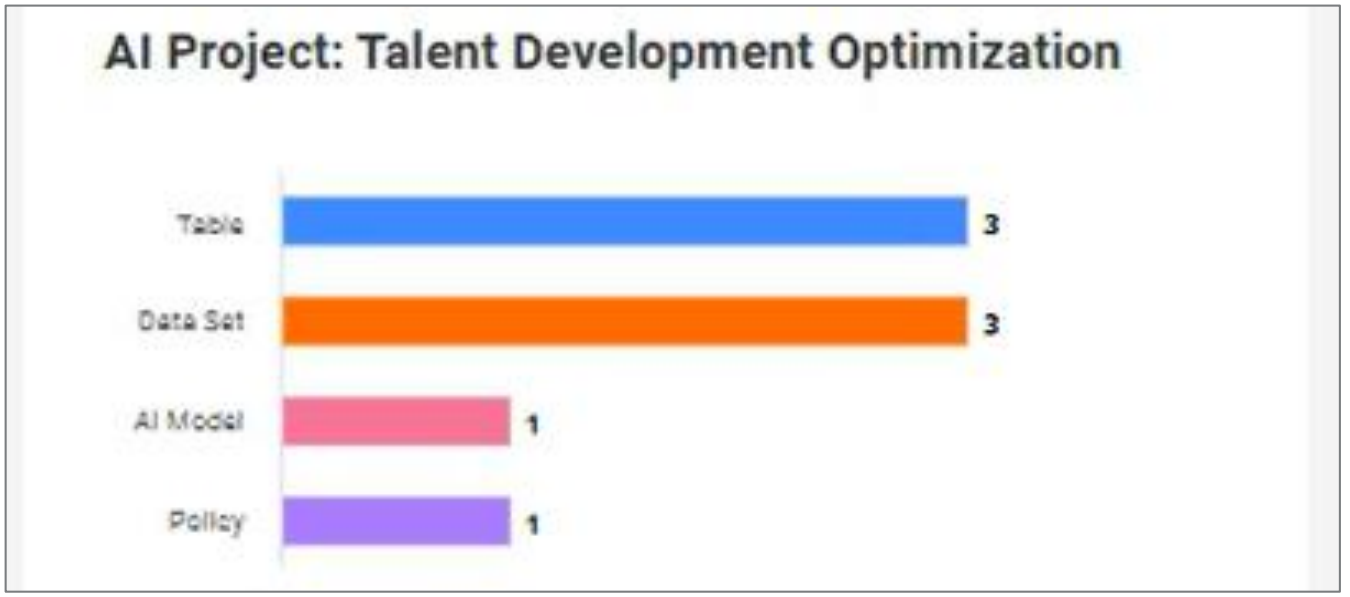
Apply Data Quality Rules Consistently Based on Business Term Level

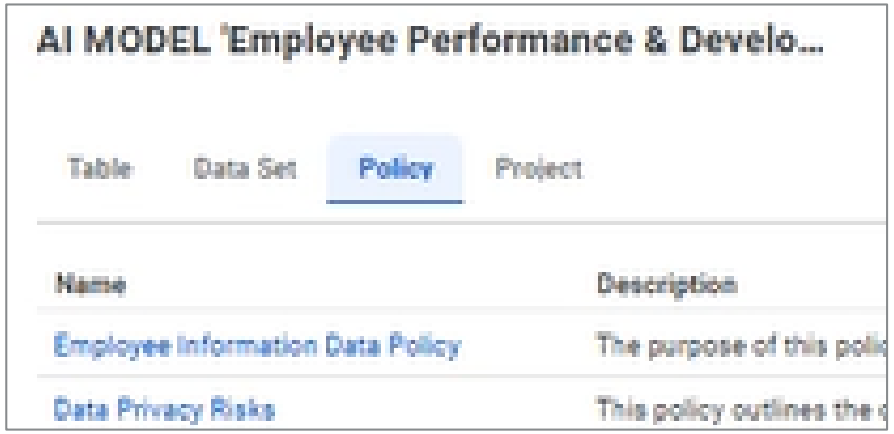
Govern the Approval Process for AI Data Usage

Enforce Controls Through an Enterprise Policy Metadata Store

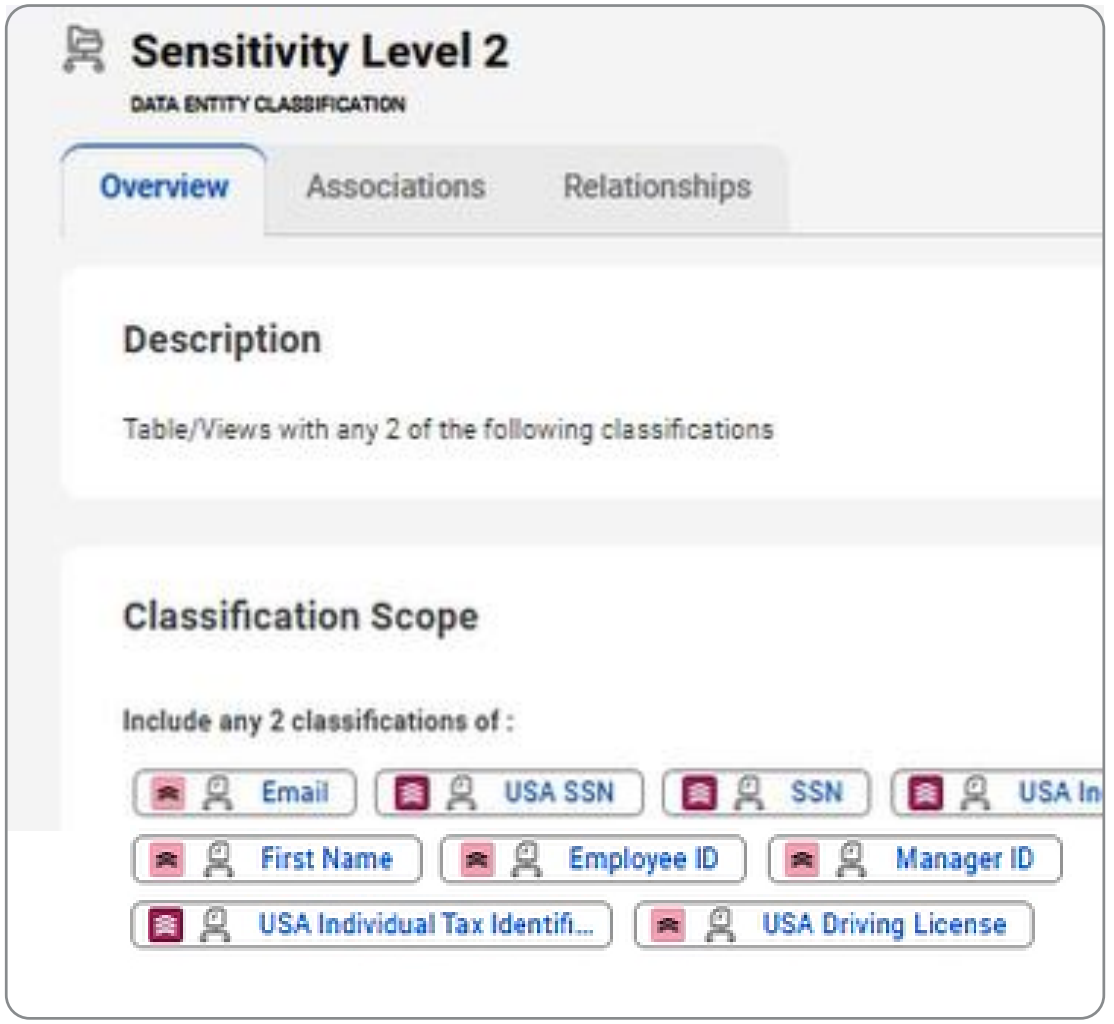
CDGC Auto-Discovers AI-Related Assets



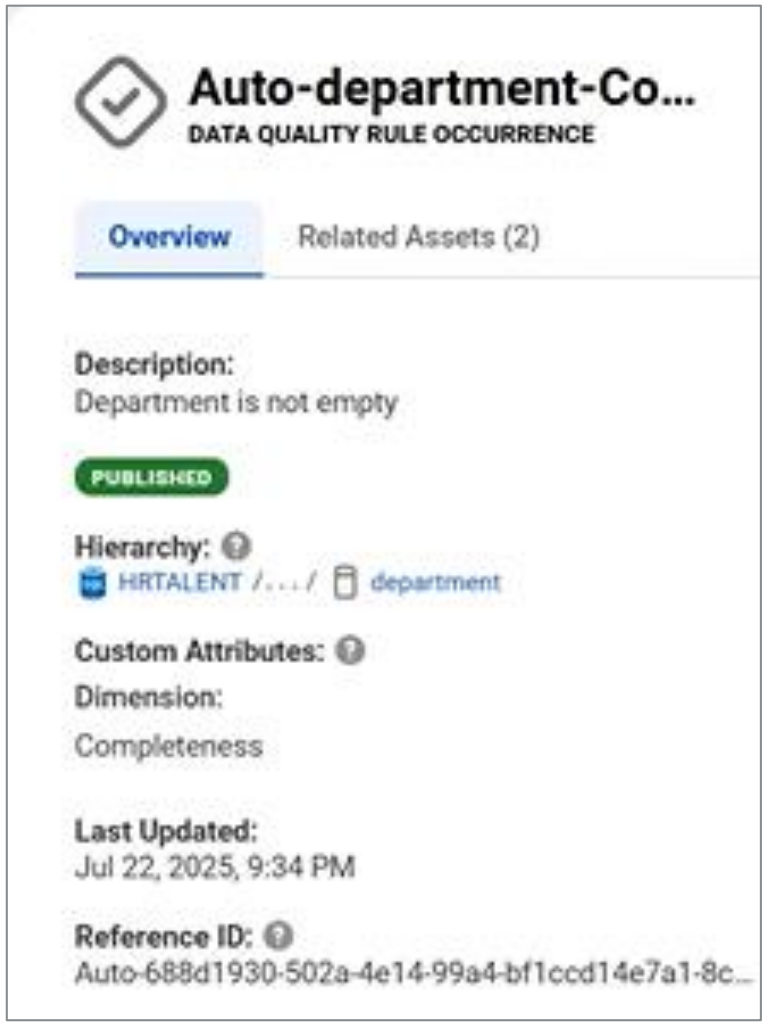


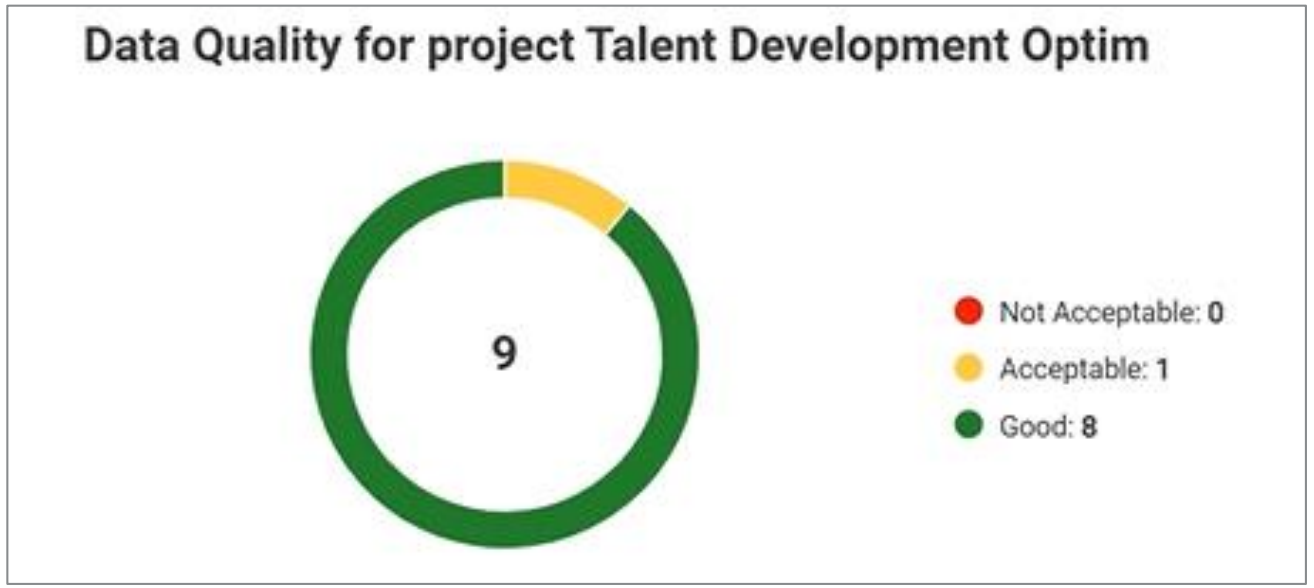


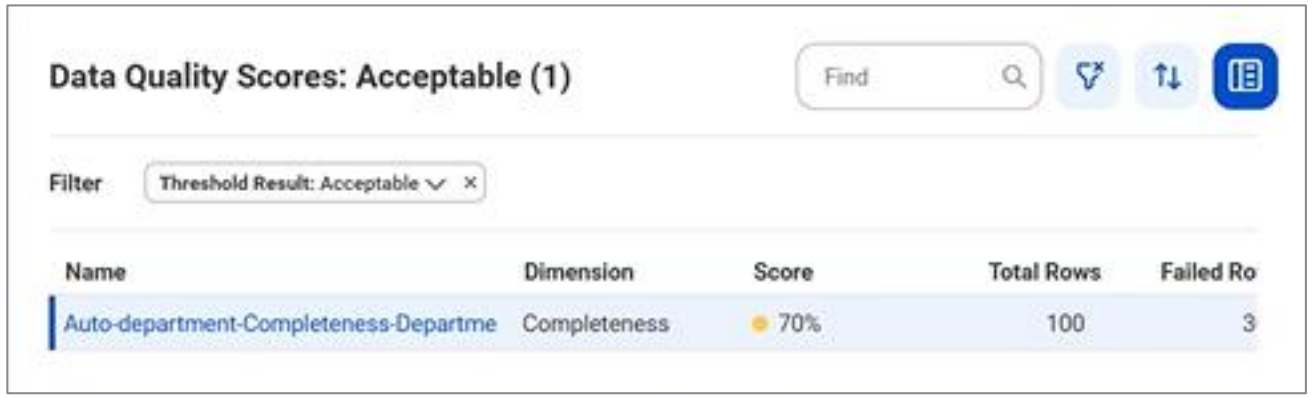
PII Discovery and Auto-Classify Highly Sensitive Data



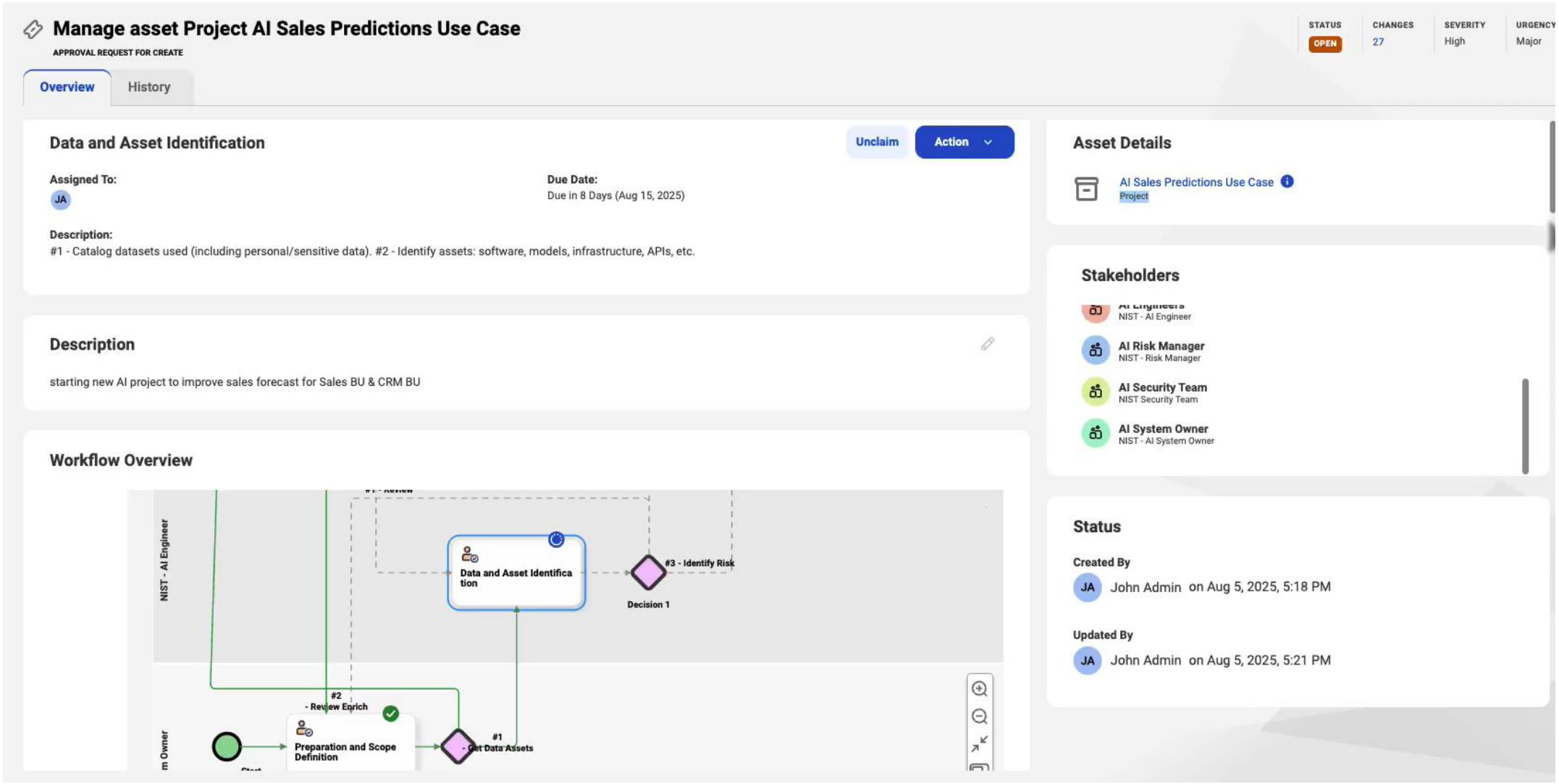
Apply DQ Rules Consistently To Business Term







Govern the Approval Process for AI Data Usage



CDGC Enhances GenAI AGILITY

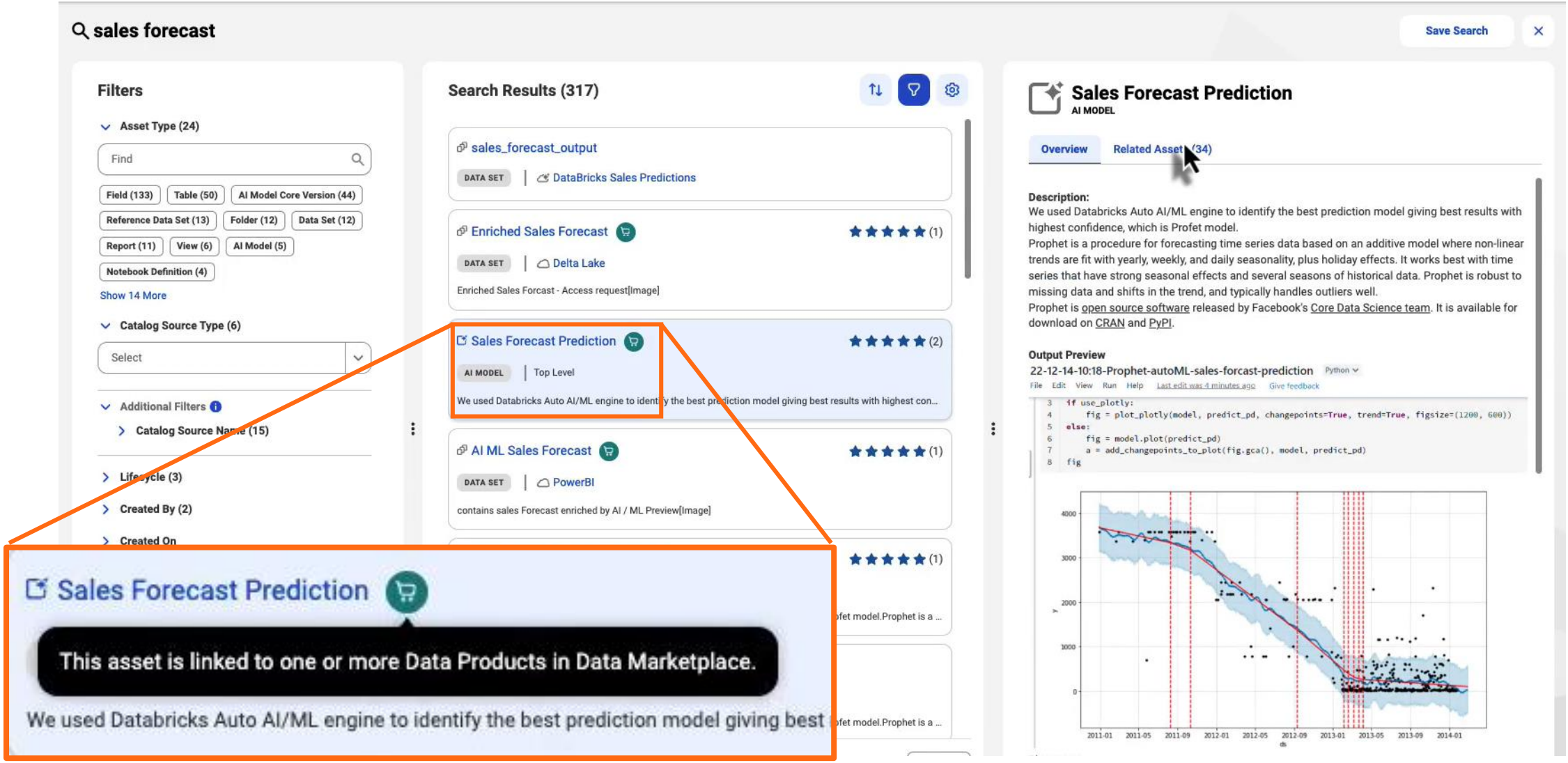
Publish Certified, AI-Ready Data Assets (Including Data Marketplace Integration)

Clarify Governance Accountability Through Workflow Timeline

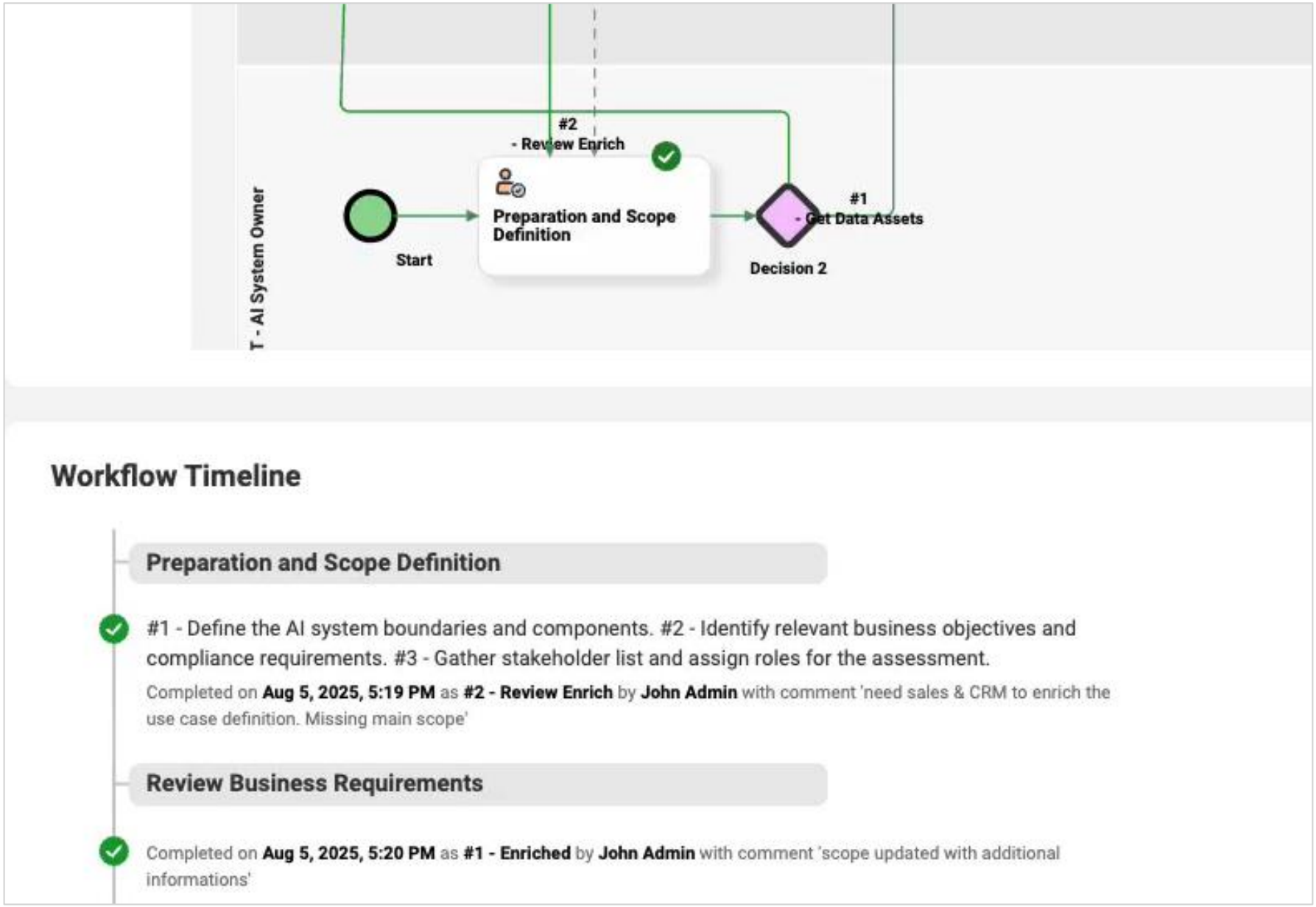
Accelerate with AI Model Lineage and Data Flow Visibility

Informatica

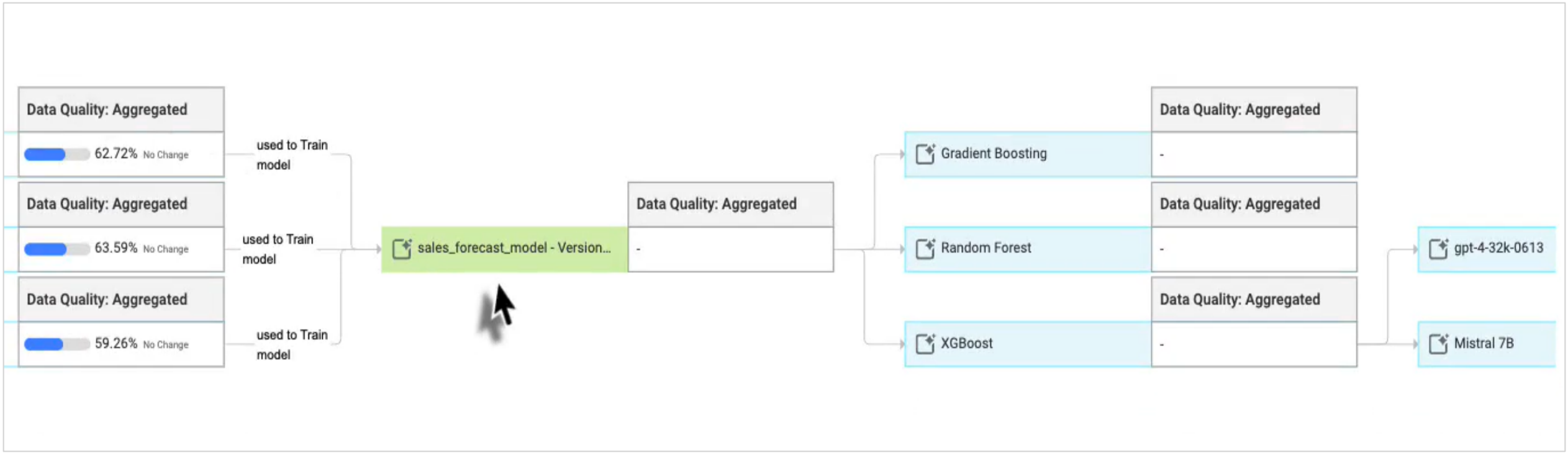
Publish Certified AI-Ready Assets to Marketplace



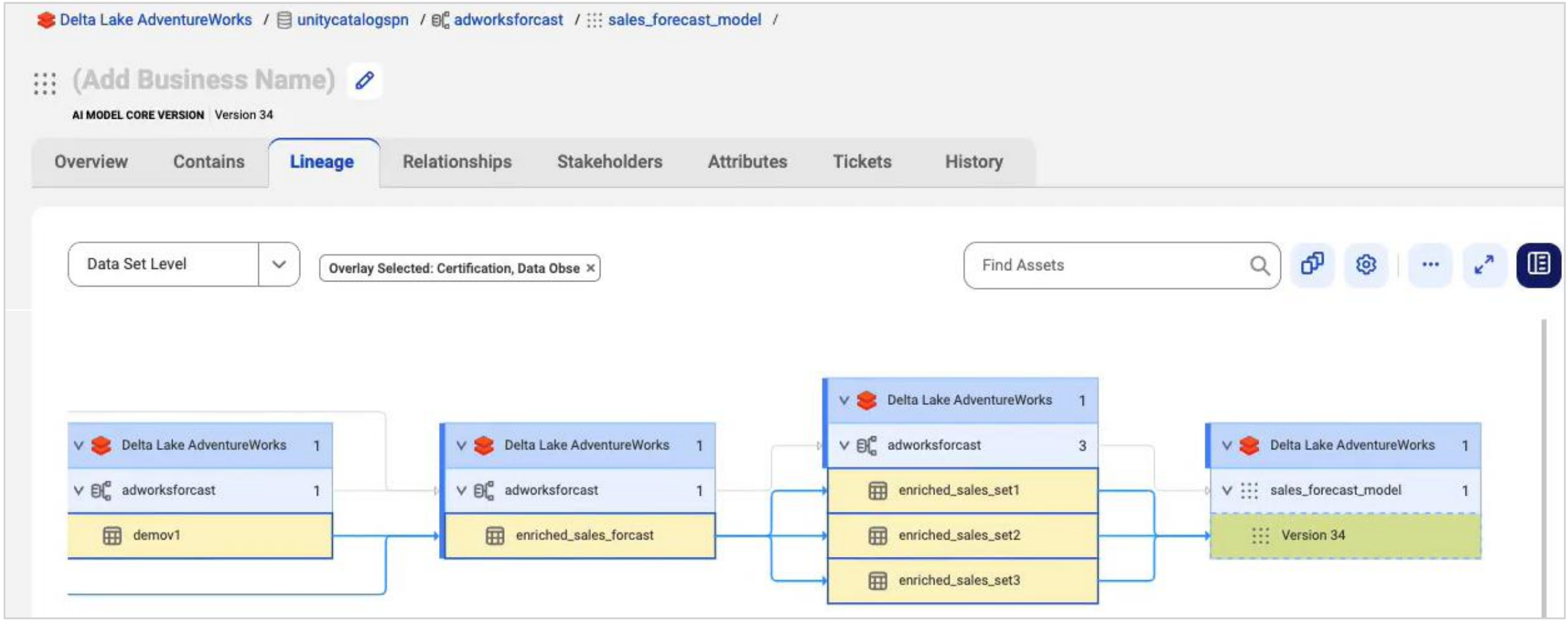
Accountability Through Data Stewardship Assignment and Audit Trail



AI Model Lineage and Data Flow Visibility



AI Model Lineage and Data Flow Visibility



Informatica

How MDM Enhances GenAI

Where data & AI come to LIFE

MDM Enhances GenAI…

- QUALITY - Reduce AI Hallucinations Through Advanced Entity Matching
  - Superior Matching Technology & AI-Augmented Algorithm
  - SSA NAME3 – Provides Cultural Context & Localization Based on Proprietary Data
  - Product Data & Attributes Consistency for AI
- RESPONSIBILITY – Provide Enterprise Single Source of Truth for Master and Reference Entities
  - Prevent AI from Consuming Unverified or Duplicate Data
  - Maintain Controlled Values Through Governed Reference Data
  - Enforce Human In The Loop in Entity Resolution
- AGILITY – Assure Trust in Master Data Across-Systems
  - Golden Record Survivorship Rule Provides Full Traceability and Trust
  - Synchronization Across Systems with Change Data Capture
  - Match Tuning Process For Ongoing Reliability

MDM Enhances GenAI QUALITY

Superior Entity Resolution / Matching Technology

Cultural Context & Localization Based on Proprietary Data

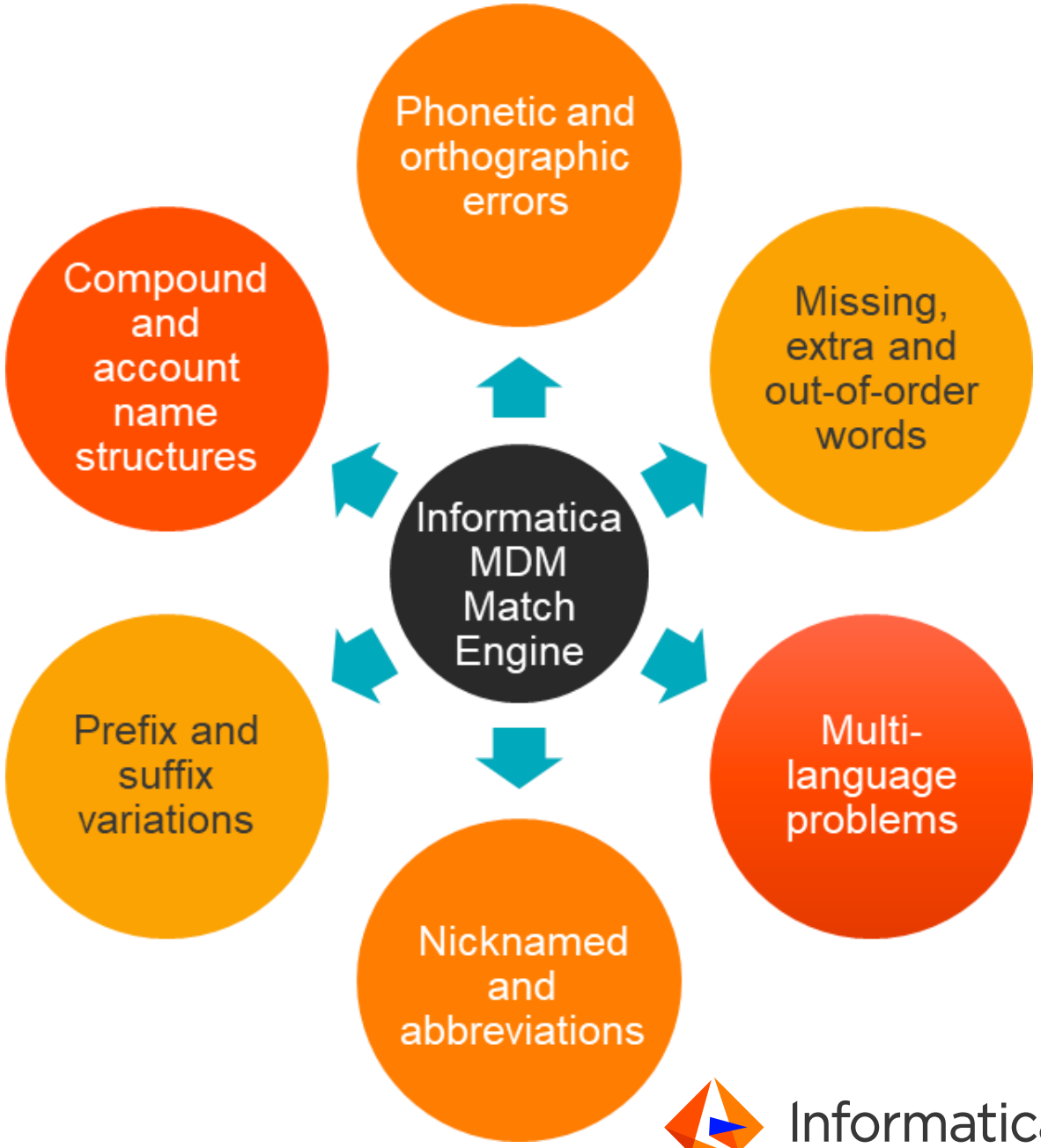
Product Data & Attributes Consistency for AI

MDM is Foundation for Intelligent 360 View of Enterprise Data

Match Technology is the Core of MDM

Informatica has the world's best matching engine – used by more than 1400 organizations, built over 30 years, covering 50+ Populations

- *Superior data matching leverages best-in-class algorithms*
- *Proven technology with more than 30 years of intelligence*
- *Out-of-the-box rules for global country populations, language and character sets*
- *Overcomes the error and variations in the data*
- *High performance for extreme data volumes*



Informatica

Cultural Context & Localization

Session	System	Population
6291456	USA_ORIG	usa

Controls

PURPOSE=Resident MATCH_LEVEL=Conservative

Response	Messages
0	

Search Data

*Person_Name*Bill Rojo*Address_Part1*436 N Forest Ave***

File Data

*Person_Name*Vilmos Rojo*Address_Part1*436 N Forest Ave***

☐ Hex

	Decision	Score
Conservative	A	097
Typical	A	098
Loose	A	099

SSA NAME3 Proprietary Data

WILLIAM
WILLAHELM
VILMOS
WILAM
HELMI
UILLEAM
GWYLLIM
WM
GUILHERME
VILPPU
WIELHELM
UILLIAM
WIM
WLM
HELM
VILEM
VILJO
WILH
WILL
VILH

Proprietary Identity Data vs. Open-Source Algorithms

Combination of proprietary data, methods, and algorithms for different classes of error and variation in identity data.

Entity Resolution Task	Example	Handled by Typical NLP Libraries*	Need Special Match / Proprietary DB**
Standard Entity Recognition	"Microsoft" as Organization	Yes	Yes
Nicknames & Aliases	“Jack" for “John"	No	Yes
Sequence Errors	"Mark Douglas" vs. "Douglas Mark"	Limited	Yes
Concatenation	"Mary Anne" vs. "Maryanne"	Limited	Yes
Localization / Cultural Variations	"Stanislav Milosovich" vs. "Stan Milo"	No	Yes
Phonetic Variations	"Edinburgh" vs. "Edinborough"	No	Yes
Transliteration / Romanization	"山田太郎" vs. "Taro Yamada"	No	Yes

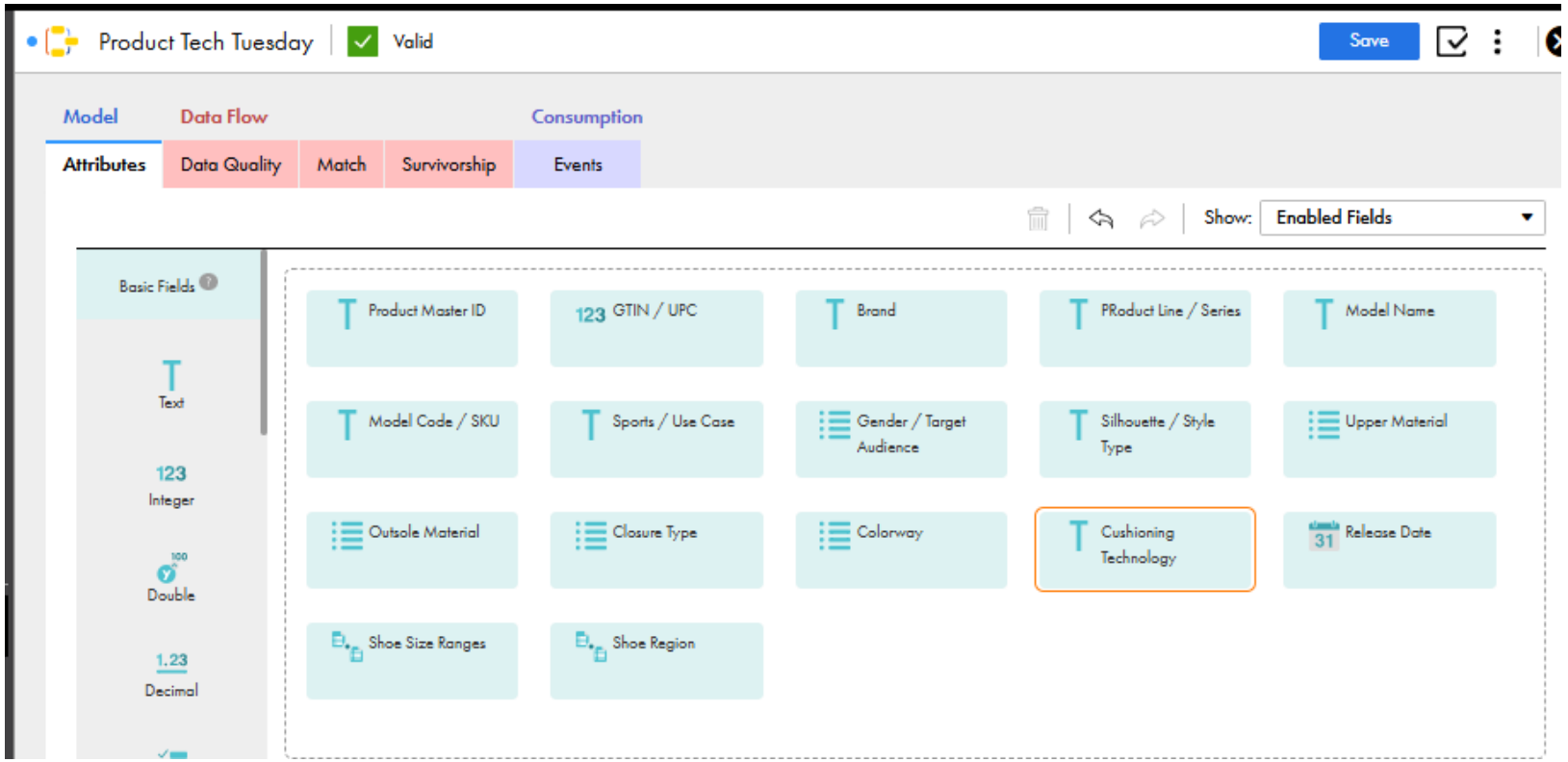
*** NLP Libraries (e.g., SpaCy):**

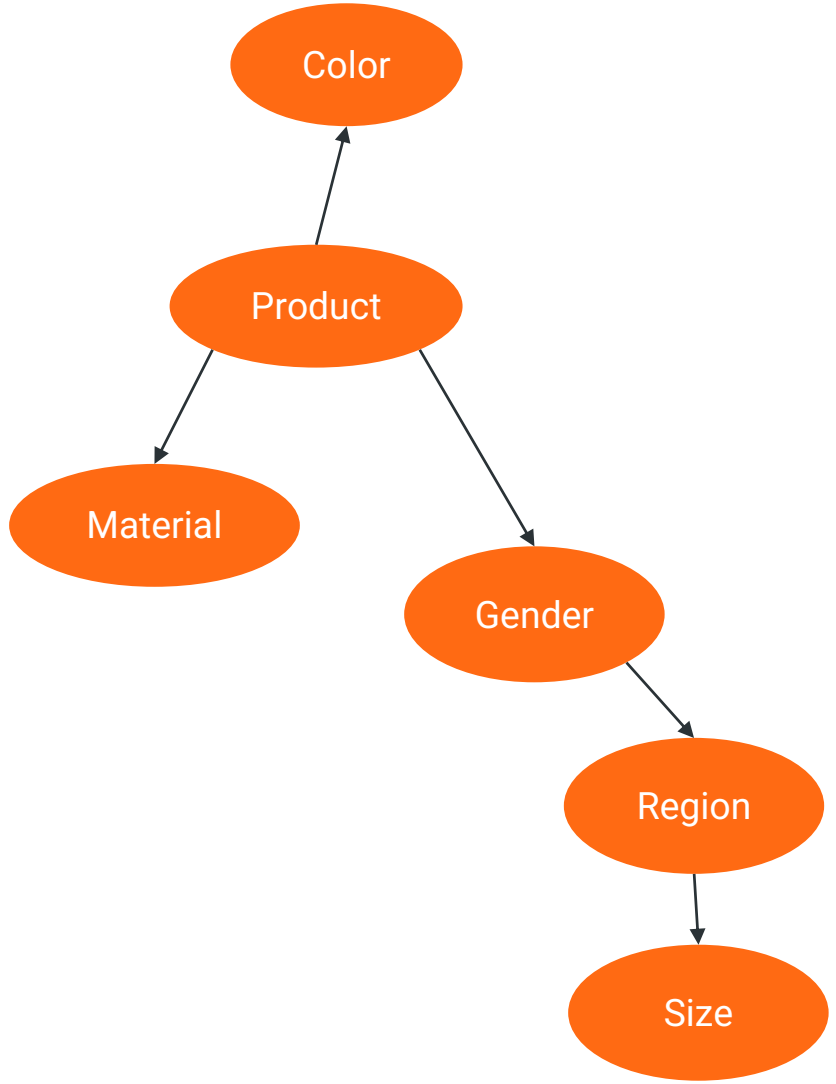
- Use statistical models and linguistic patterns
- Handle many standard errors
- Limited with custom errors like nicknames, localization, or rare abbreviations

**** Informatica’s SSA-NAME3 DB:**

- Proprietary databases w advanced match algorithms
- Handle complex, domain-specific variations and cultural differences
- Deliver higher accuracy through curated knowledge

Product Master Data Example





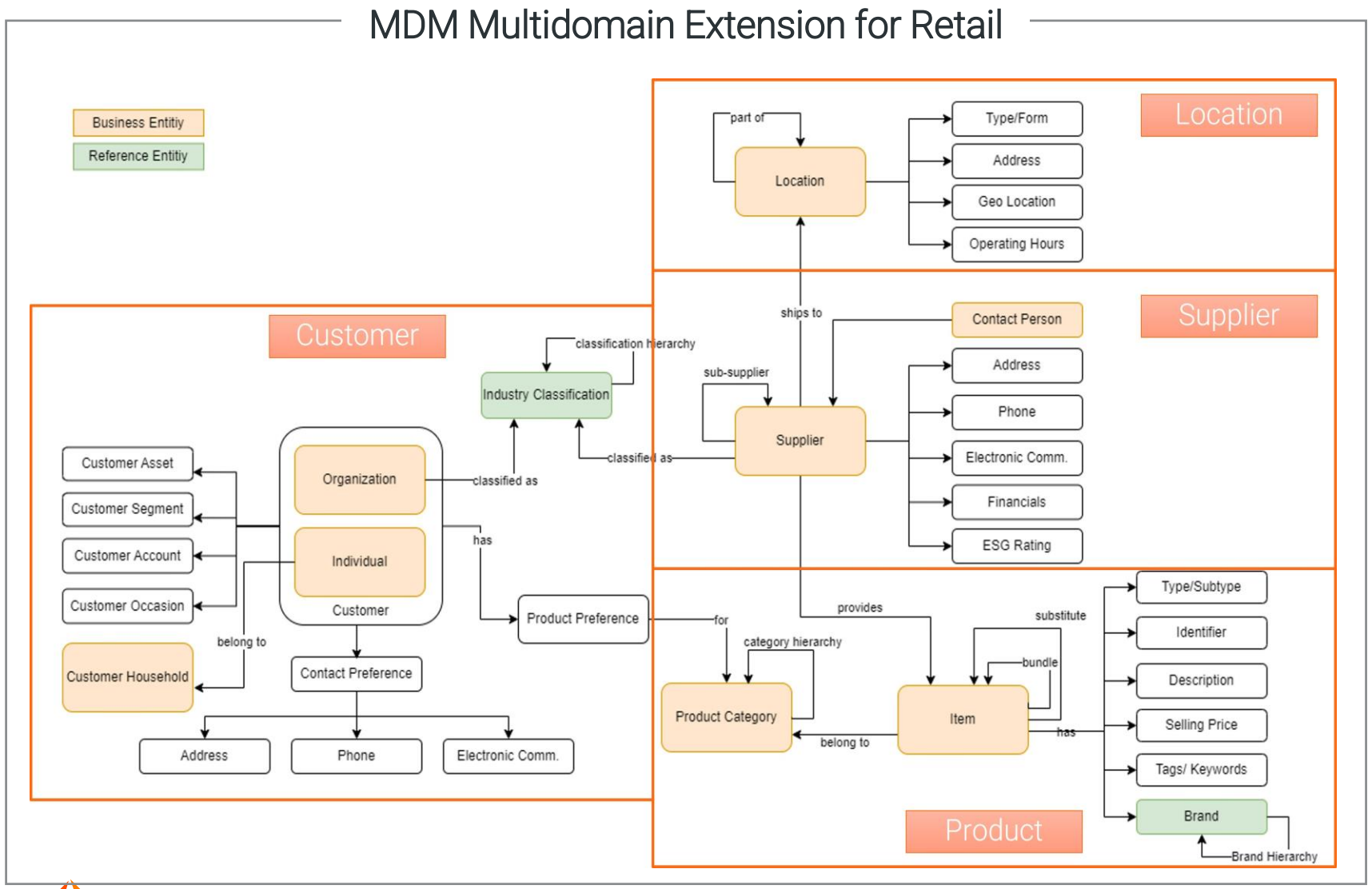
MDM Enhances GenAI RESPONSIBILITY

Ground GenAI on Authoritative Business & Metadata Model

Govern Regulatory-Sensitive Values Through Reference Data

Enforce Human In The Loop in Entity Resolution

The ONLY True Multi-Domain (140+ Domains)



The ONLY
True Multi-Domain
140+ domains

B2C Customer

B2B Customer

Product

Reference

Supplier

Location

What is Reference Data and Why It Is Critical?

Enterprise Reference Data

Origins

WHAT is it?
- Lookup Values
- Reference Codes
- Reference Hierarchies

HOW is it known?
- List of Values
- Code Tables
- Needs & Desires

WHO creates?
- Data Governance
- Business Owners
- External Agencies

WHAT else?
- Hierarchical
- Relatively Static
- Multiple Versions
- Multiple Cardinality

Uses

WHERE?

Master data
- Customer Data
- Product Data
- Location Data…

Transactional data

Application data
- CRM
- ERP…

Interaction data

Almost *any* data

Impacts

- Reduce Risk
- Improves Business Intelligence
- Compliance Reporting
- Data Quality
- Productivity
- Sale & Marketing…

Influences

Operational Efficiency

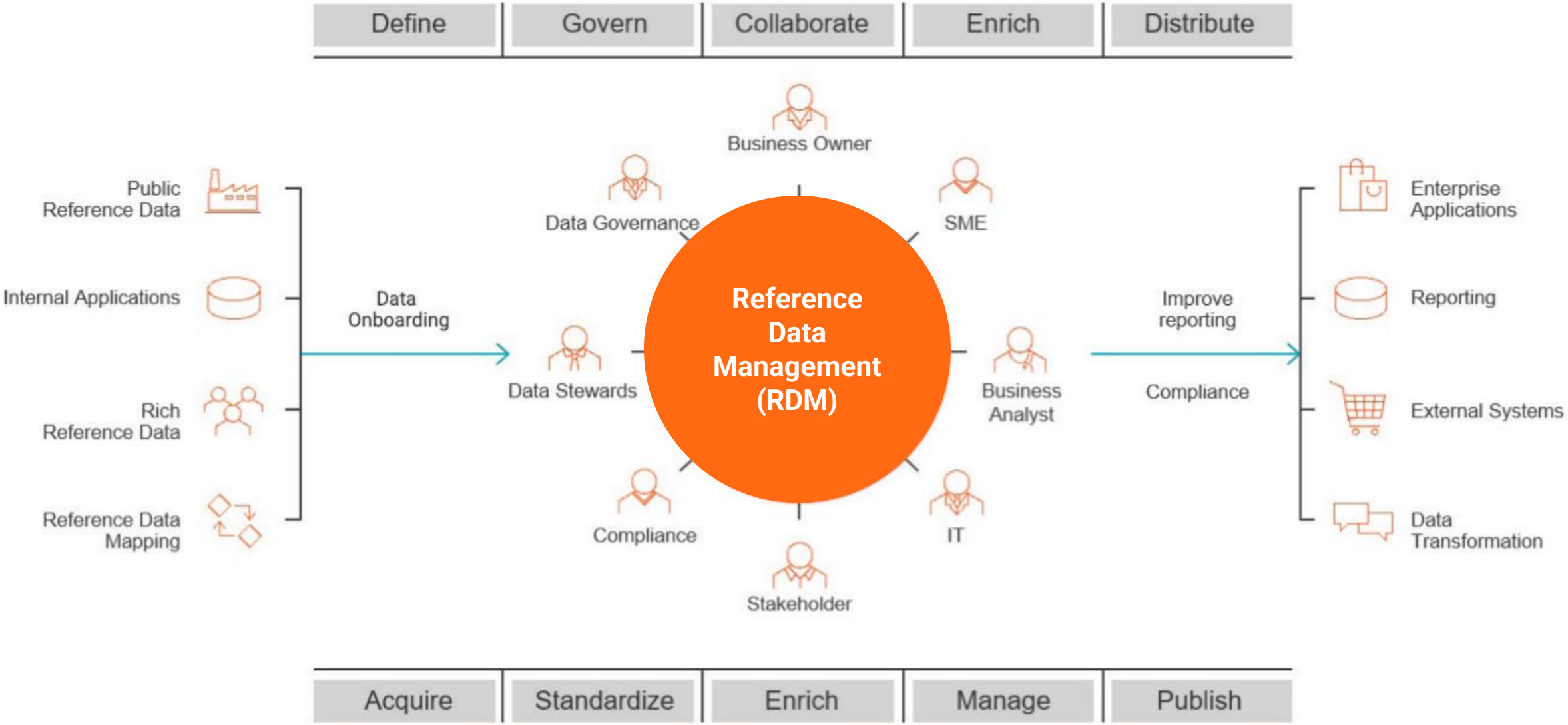
Enterprise Apps

Customer Experience

Compliance

Informatica

Govern Regulatory-Sensitive Reference Data



Product Master & Reference Data Lookup Example

Enterprise Shoe Material (5) Point in Time Current

Name ↑	Code	Description
Engineered Mesh	ENG_MSH	Lightweight, breathable mesh fa
Knit Fabric	KNIT	Flexible, sock-like upper
Natural Leather	NAT_LEA	Genuine leather
Rubber	RUB	Outsole material for traction
Synthetic Leather	SYN_LEA	Durable leather substitute

Enterprise Shoe Color (6) Point in Time Current

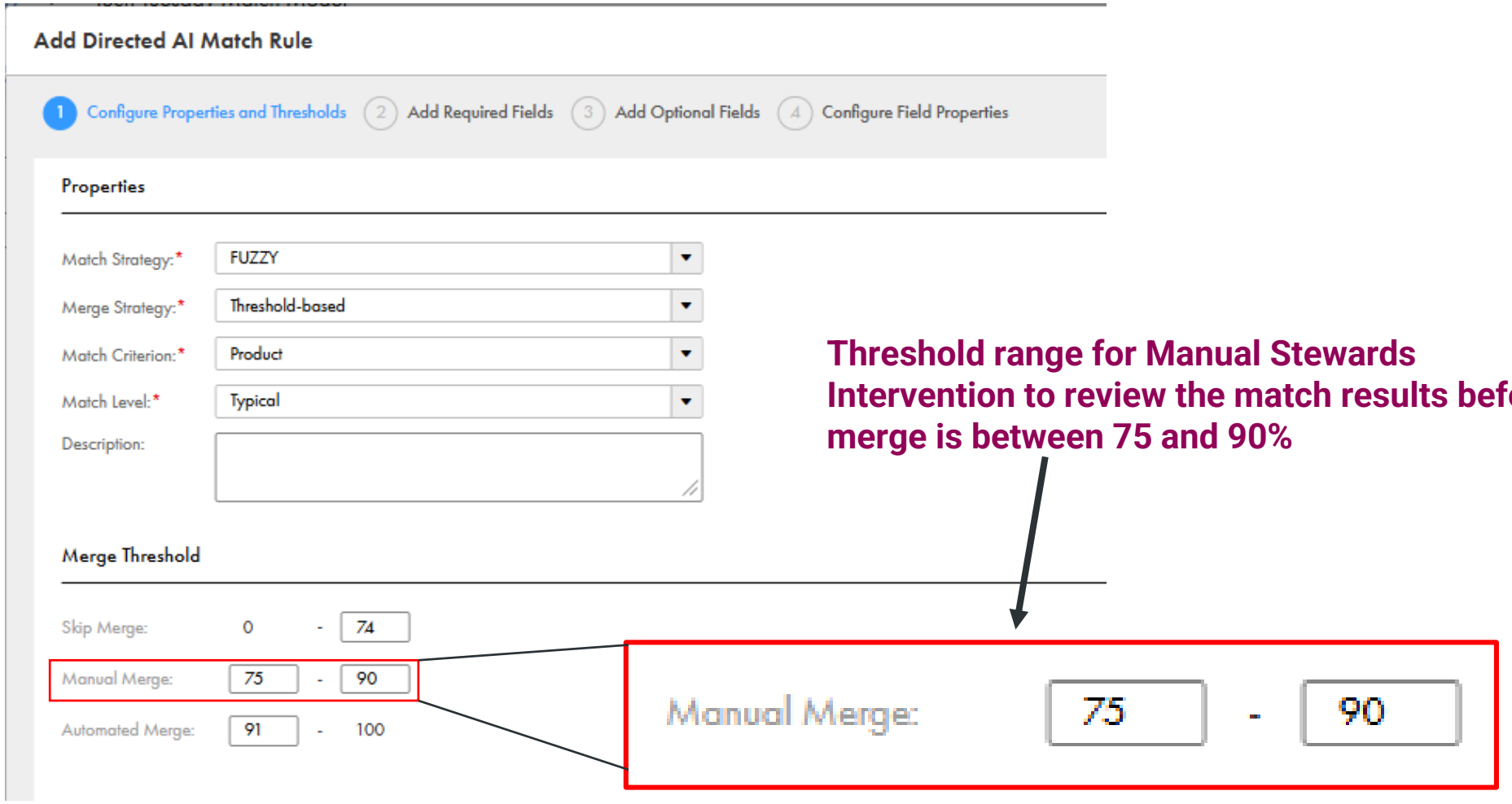
Name ↑	Code	Description
Black	BLK	Black, Jet Black, Onyx
Blue	BLU	Blue, Navy, Royal Blue
Crimson	CRIM	Crimson, Deep Red, Ruby
Gray	GRY	Gray, Ash, Charcoal
Orange	ORG	Orange, Tangerine, Blaze
White	WHT	White, Snow, Pearl

Enterprise TT Shoe Size v1 (12) Point in Time Current

Name ↑	Code	Description	dependency
UK Female 8.5	UK_FEMALE_85	UK Women 8.5	UK Female
UK Female 9.0	UK_FEMALE_9	UK Women 9.0	UK Female
UK Female 9.5	UK_FEMALE_95	UK Women 9.5	UK Female
UK Male 8	UK_MALE_8	UK Men 8	UK Male
UK Male 8.5	UK_MALE_85	UK Male 8.5	UK Male
UK Male 9	UK_MALE_9	UK Men 9	UK Male
US Female 10	US_FEMALE_10	US Women 10	US Female
US Female 10.5	US_FEMALE_105	US Women 10.5	US Female
US Female 11	US_FEMALE_11	US Women 11	US Female
US MALE 9	US_MALE_9	US Men 9	US Male
US Male 8.5	US_MALE_85	US Men 8.5	US Male
US Male 9.5	US_MALE_95	US Men 9.5	US Male

Informatica

Enforce Human In The Loop in Entity Resolution



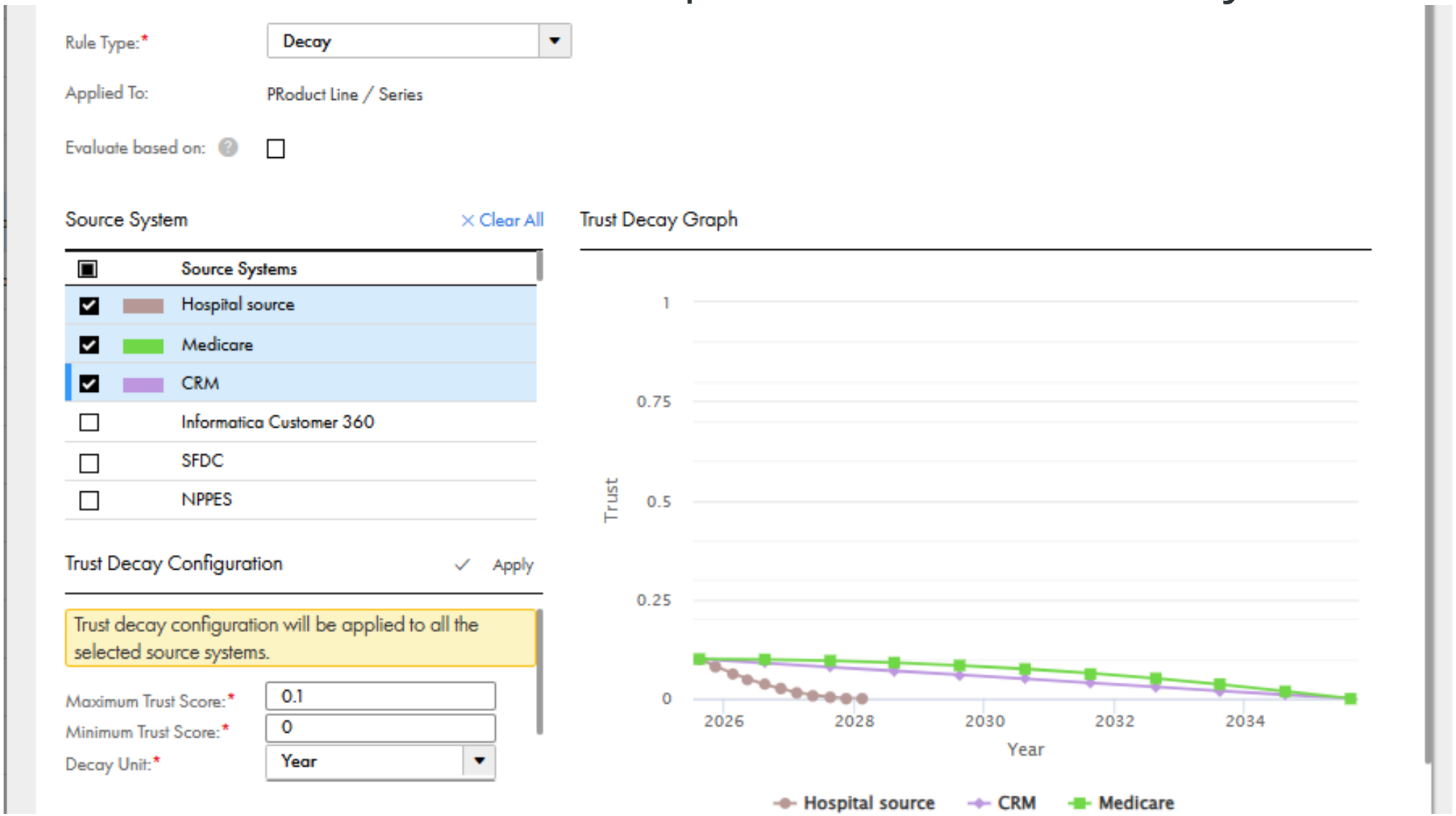
MDM Enhances GenAI AGILITY

Golden Record Survivorship Rule Provides Full Traceability and Trust

Synchronization Across Systems with Change Data Capture

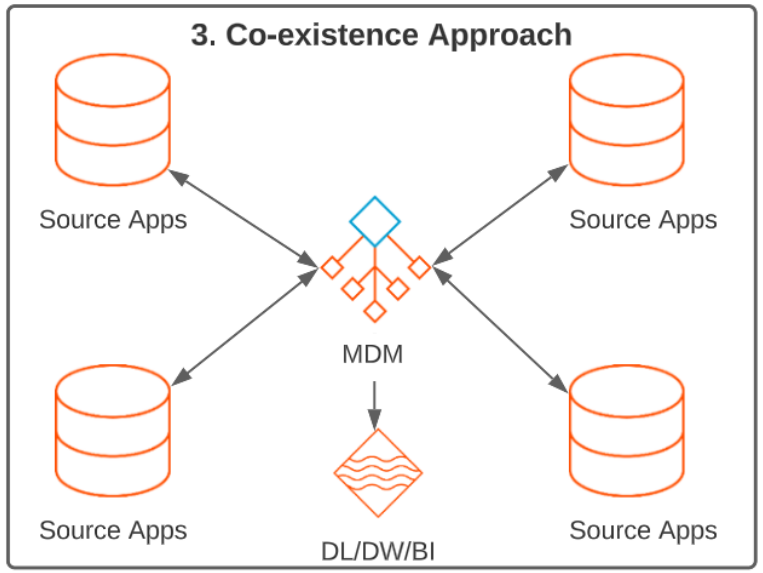
Match Tuning Process For Ongoing Reliability

Golden Record Survivorship - Full Traceability and Trust



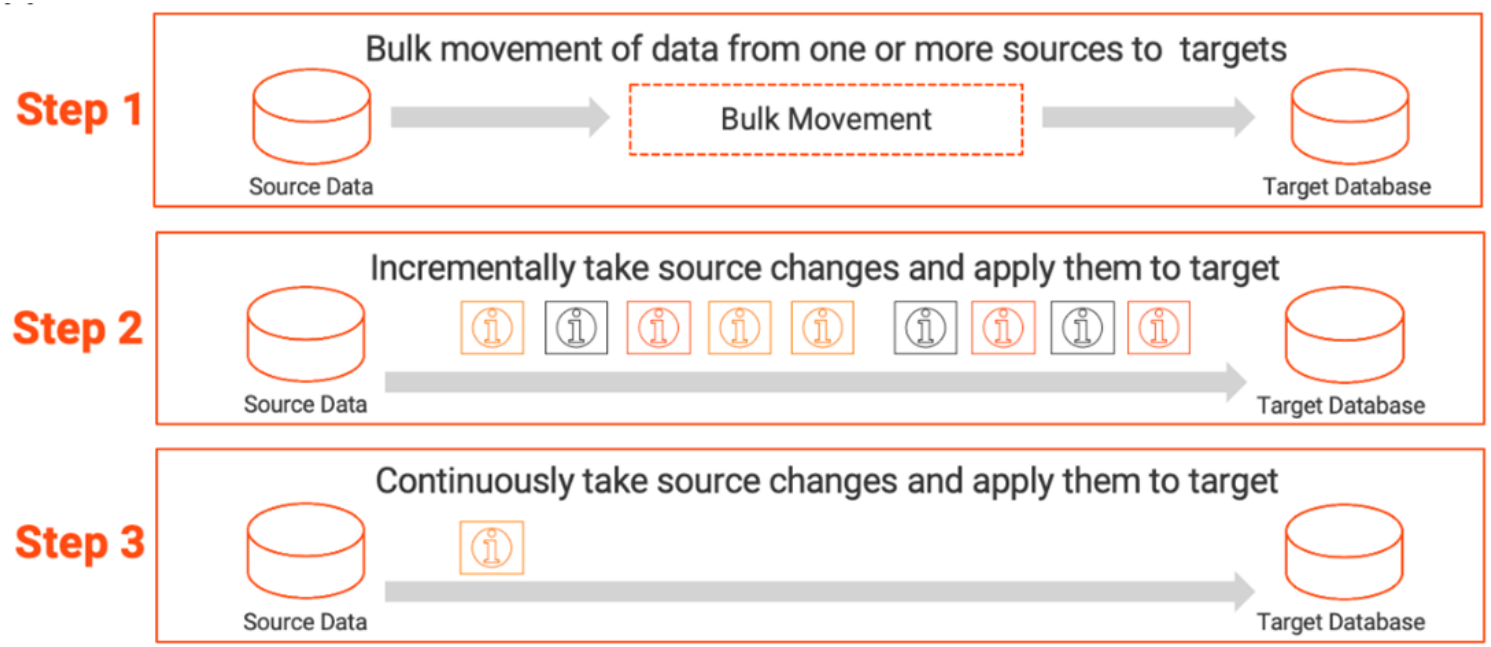
Synchronization of Golden Values Across Systems

Publishing Changes Across Sources



- Credit to Gartner for proposing this set of approaches

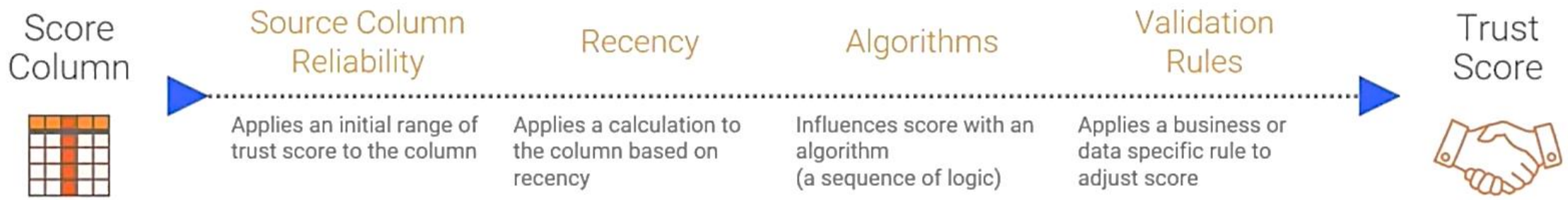
Push Changes in either Bulk, Incremental, or Continuous



Superior Match Algorithm Tuning with Trust Framework

Reliability	Recency	Data Decay	Rules	Over-Rides	Intelligence
Reliability of the source data	Reliability of the source data	What kind of data is it?	Validation rules to downgrade trust score	Manual adjustment after the fact	Weighted system to assign trust scores

Informatica Trust Framework

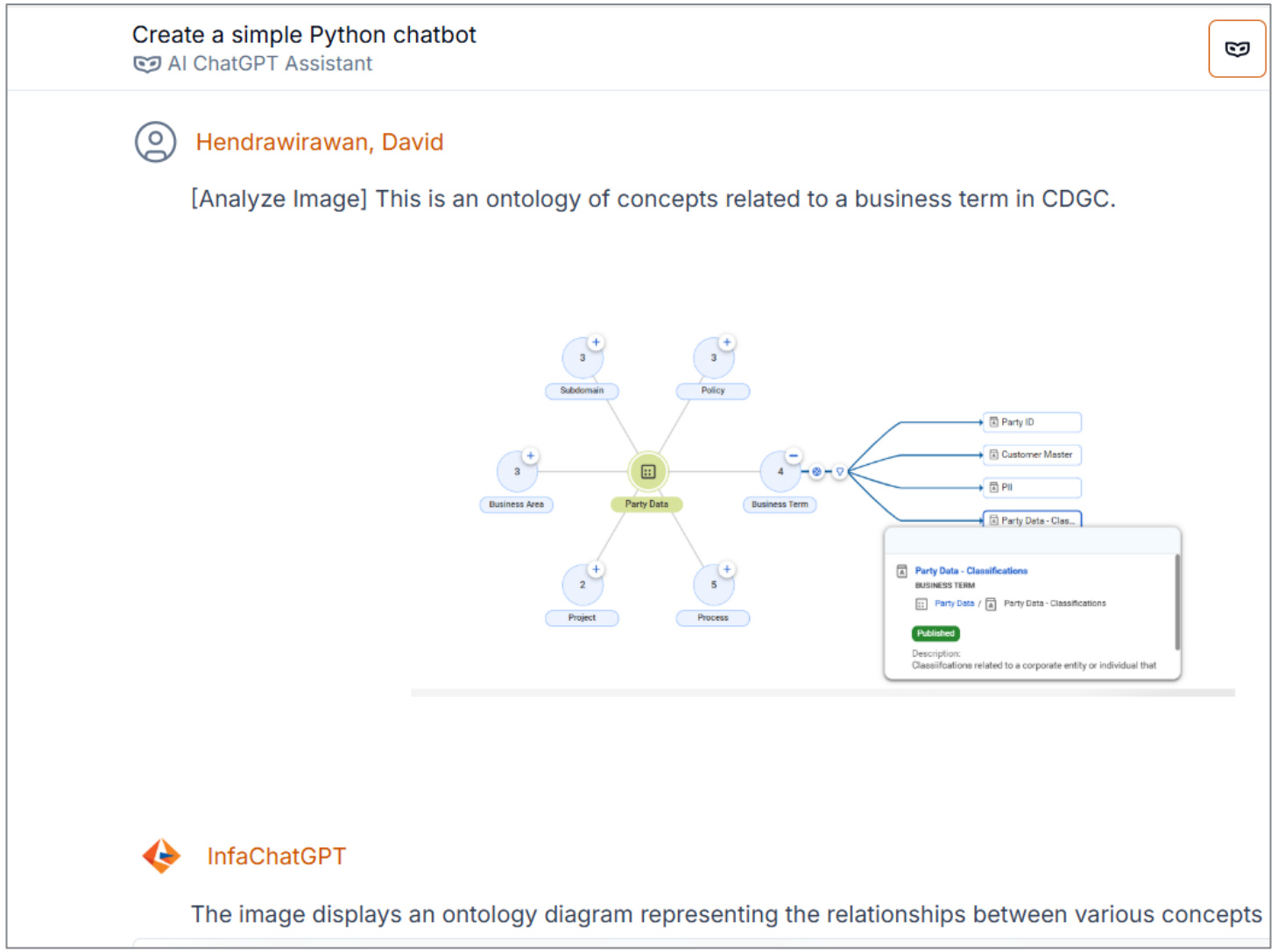


Demo: Enhancing Chatbot w/ CDGC & MDM

Prompt 1 – V1 (basic chatbot)
Create a simple Python chatbot that
- Use a small in-memory list of products
- Ask clarifying questions until it narrows to a few items
- Short and simple, no data controls

Prompt 2 – V2 (reinforced with enablers)
Now improve the chatbot by reinforcing it with enterprise data structures.
1. Reference data
2. Master data
3. Metadata/taxonomy
4. Provenance/lineage

Prompt 3 – V3 (with CDGC + MDM)
Instead of hard-coding, rewrite the chatbot to call APIs:
- CDGC for metadata, taxonomy, glossary, synonyms, and lineage.
- MDM for golden records, reference lists, and matching rules.



Q&A

Where data & AI come to LIFE