Global IDs

A Case Study

Benefits of Global IDs enabling the business

© Copyright 2022 Global IDs - Private and Confidential Information

Executive Summary

- A large financial organization was in the process of inventorying data sources, vendors and data assets
- Next logical step was to establish the Data Governance organization
- Efficient Data Governance required a platform specifically built to address data inventory, data quality and data governance questions
- Global IDs was evaluated using a key business process
  - The outcome highlighted the functionality Global IDs brings to the table that is beneficial across several aspects of data governance:

Transparency	Traceability	Trust
• Centralized business term repository across the organization • Centralized data inventory • Support for data migration initiatives	• Centralized data view: profiling, mapping and quality assessment • Centralized view of the data flows between applications	• Data quality monitoring, including datasets delivered by 3rd parties • Outbound data quality monitoring • Data Quality issue resolution workflow

© Copyright 2022 Global IDs - Private and Confidential Information

A Case for One-Company-Data

- The company was in need to take control of its data assets
- A company-wide Data Governance Program with appropriate tools was needed
- Legacy “siloed” business and technology philosophy clearly manifests itself in process and data
  - Discrepancies in business terminology
  - Multiple duplicative databases and data sources
  - Unknown or non-existing data flow documentation
  - Unknown and undocumented data lineage
  - Unknown/undocumented data consumers and usage patterns
- High cost of data operations
- Lost revenue opportunities due to unreliable information about the data

A data governance framework with a single set of rules and processes for collecting, storing, and using data

The Data Governance program should address:

- People
- Processes
- Contributors (to Data Pipeline)
- Technology

A Case for One-Company Global IDs Adoption

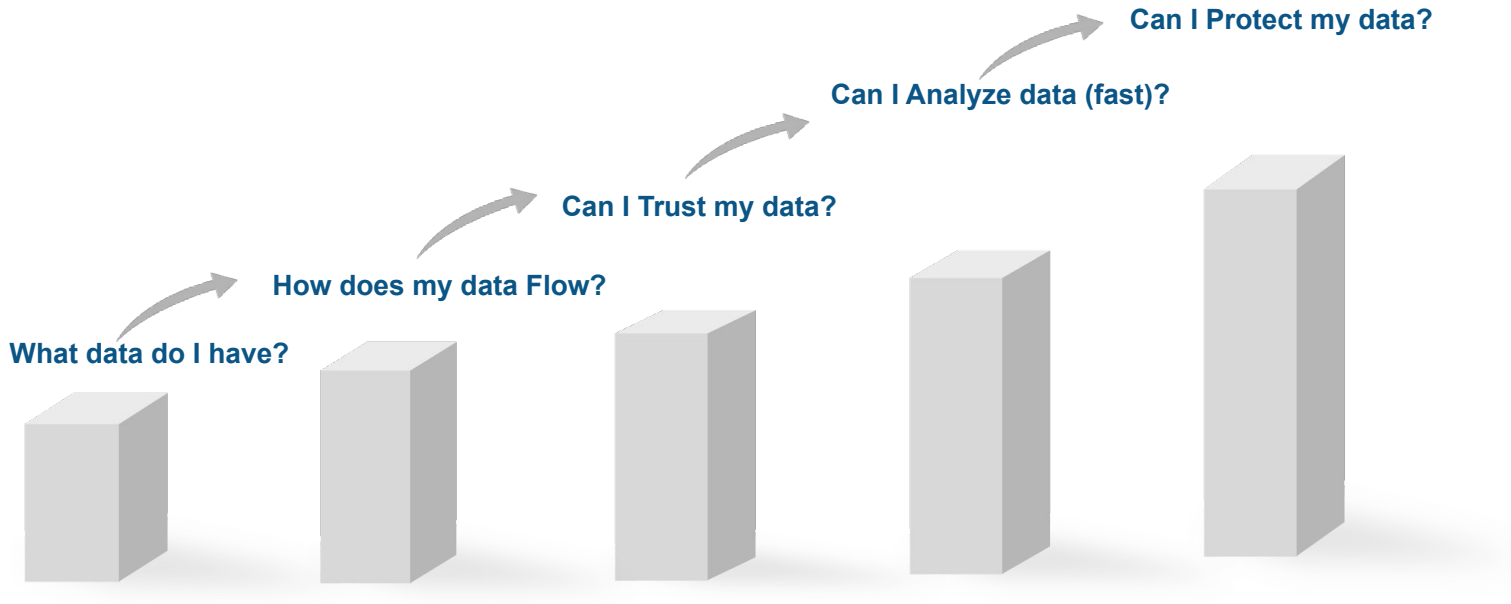
- Global IDs is a platform that is built to help organizations regain control of and gain trust in their data assets
- Facilitates use of Common Business Terminology
- Facilitates disciplined approach to Data Governance
- Goes straight to the source – Global IDs directly connects to databases and evaluates real data
- Facilitates all aspects of Data Governance:
  - Transparency (Where is the data?)
  - Traceability (How is data flowing?)
  - Trust (How reliable is data?)
  - Analytics (Are we deriving value from data?)
  - Privacy (Is sensitive data protected?)
- Ultimately reduces cost of data operations
- Opens new revenue opportunities by providing answers to business data-related questions quickly and with authority

A data governance framework with a single set of rules and processes for collecting, storing, and using data.

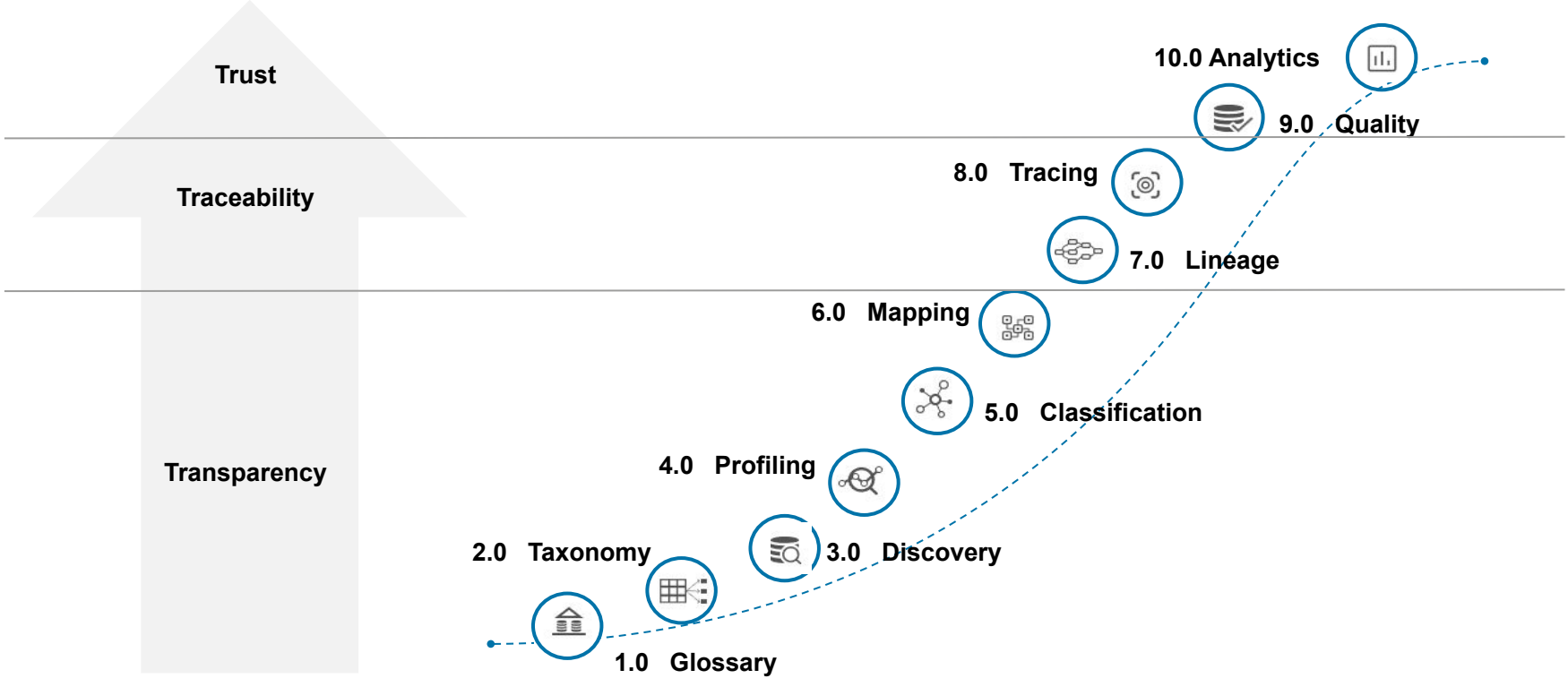
The Data Governance program should address:

- People
- Processes
- Contributors (to Data Pipeline)
- Technology

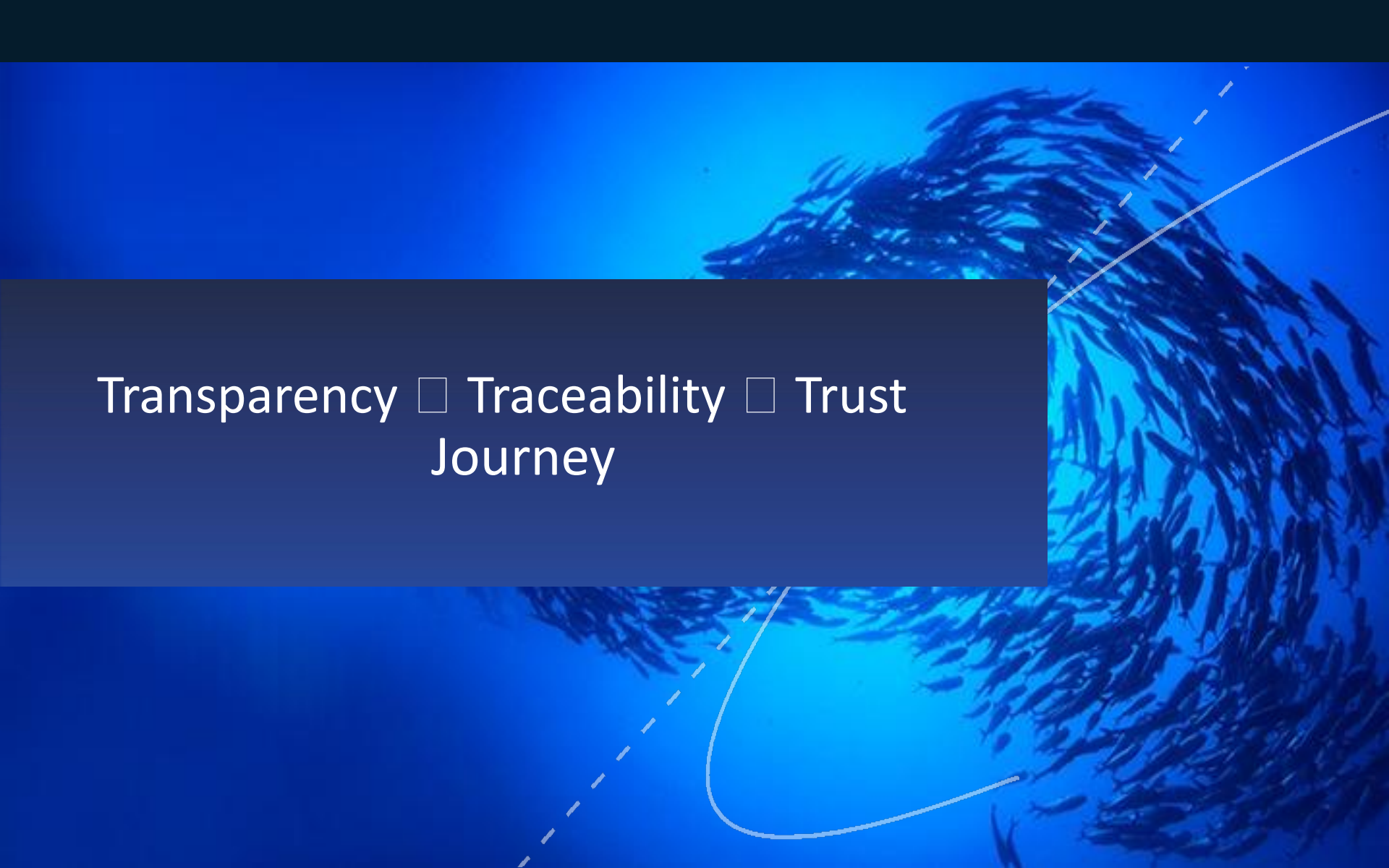
The Company needed to answer key data questions that could not be answered previously.



Global IDs Data Governance methodology is a 10 step process enabling value

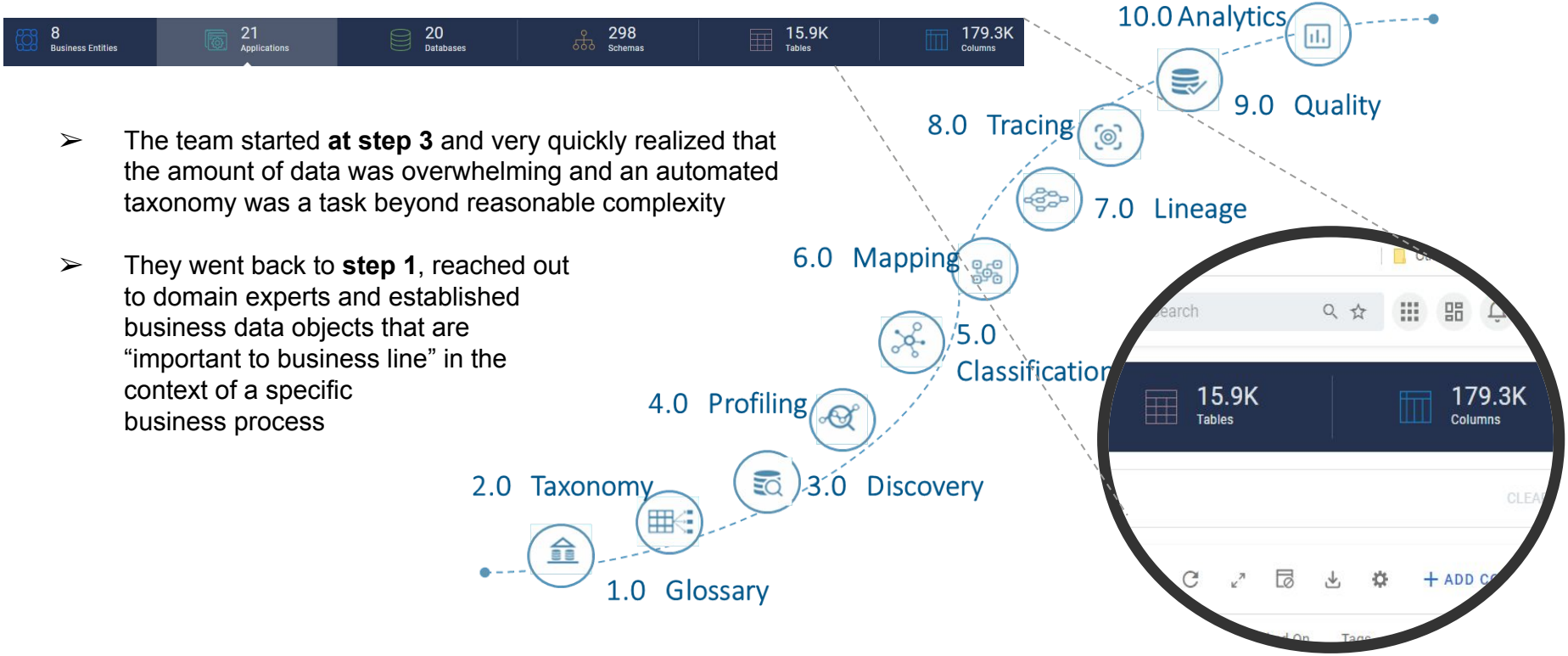


© Copyright 2022 Global IDs - Private and Confidential Information


Transparency □ Traceability □ Trust
Journey

Data Discovery Journey

- The team started **at step 3** and very quickly realized that the amount of data was overwhelming and an automated taxonomy was a task beyond reasonable complexity
- They went back to **step 1**, reached out to domain experts and established business data objects that are “important to business line” in the context of a specific business process



© Copyright 2022 Global IDs - Private and Confidential Information

The importance of speaking the same language

Scenario

- ➔ Lack of a common business ontology
- ➔ Lack of a common business ontology documented (neither a system to maintain one)
- ➔ Discrepancies in business terminology between business units

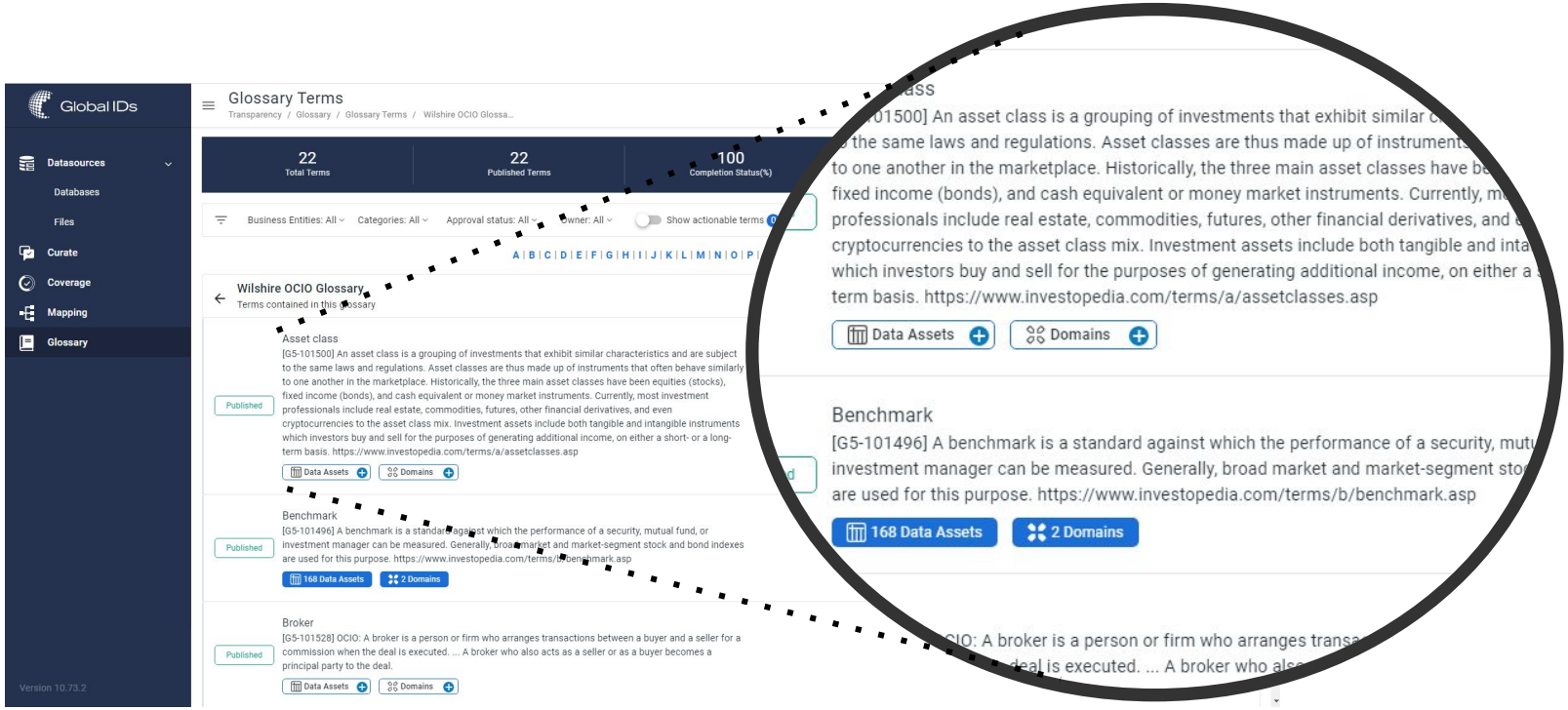
Global IDs

- ★ Glossary functionality allows one or multiple Business Terminology Glossaries with agreed upon definitions.
- ★ Glossaries are integrated and can be used to create metadata entities in the system
- ★ Glossaries can be imported
  - Public domain/industry standards (eg: FIBO)
  - Company-specific (Excel, etc.)

TRANSPARENCY

Example – Business Glossary Term

1.0 Glossary



TRANSPARENCY

Systematic representation of business terms

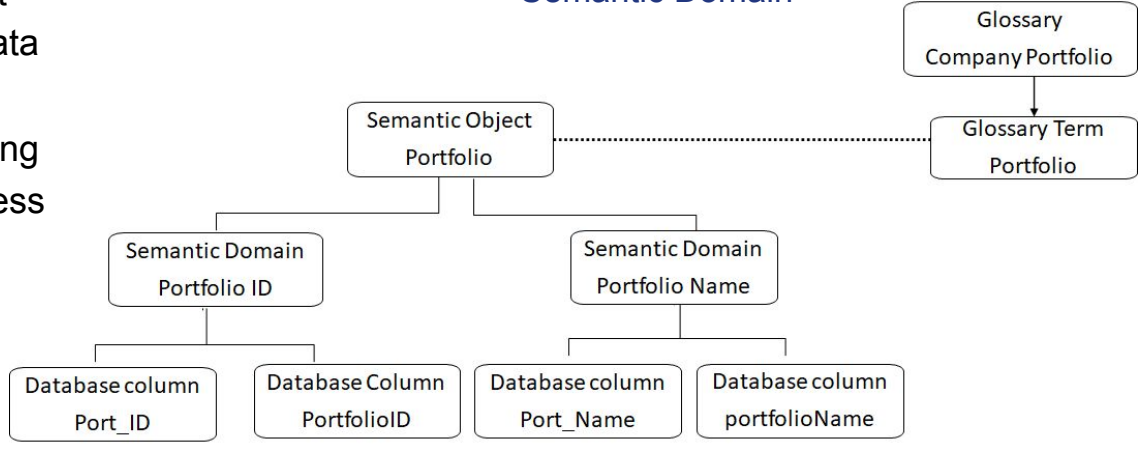
2.0 Taxonomy

Scenario

- ➔ Created in isolation, databases do not share the same “technical terminology” discipline when it comes to naming table and data elements
- ➔ There is no conceptual mapping of data elements to the business domains/terms

Global IDs

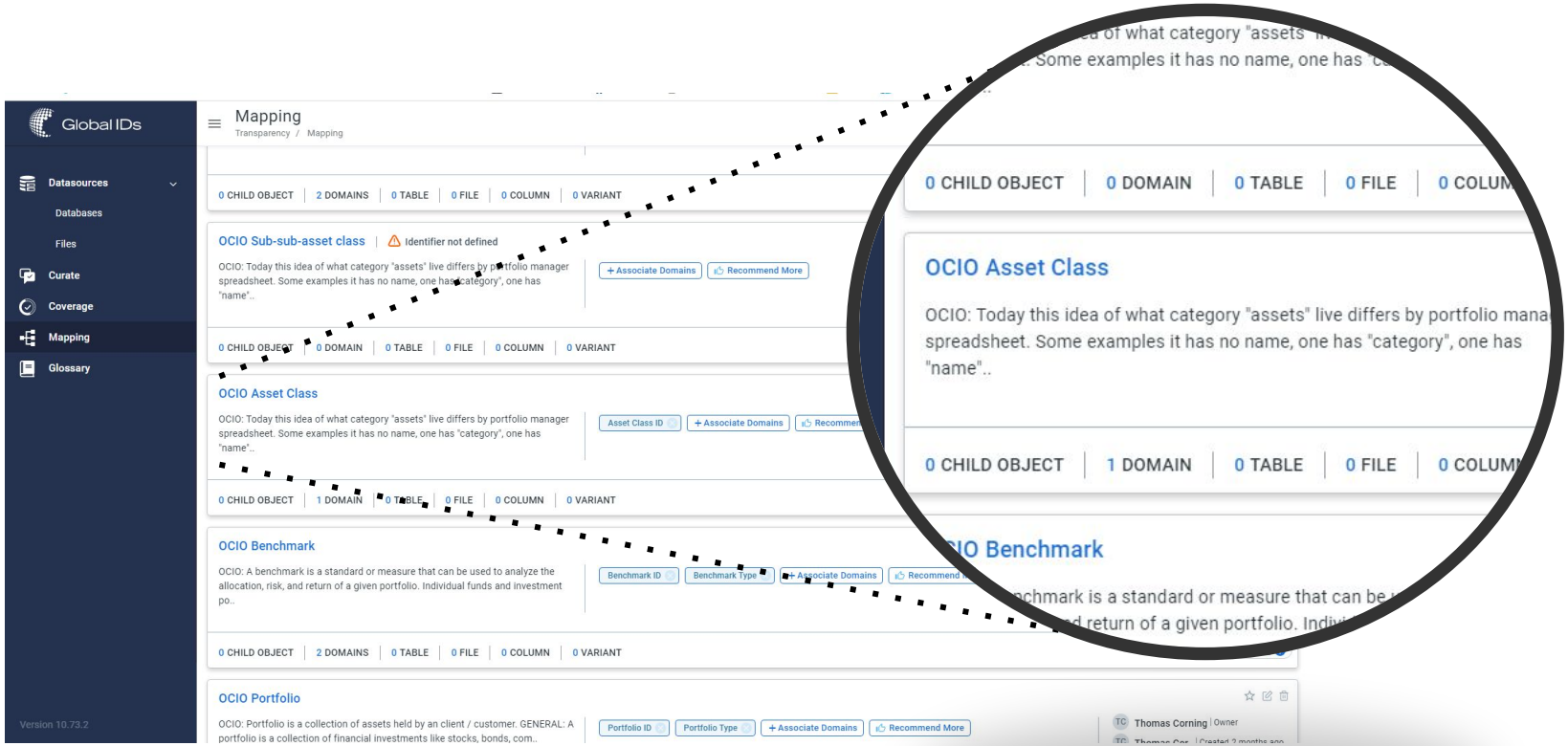
- ★ Global IDs introduces concepts of
  - Semantic Object
  - Semantic Domain



TRANSPARENCY

Example – Company Semantic Objects

2.0 Taxonomy



TRANSPARENCY

Looking into the real databases

3.0 Discovery

Scenario

- Organized a DBE working group with broad business entity and process coverage:
  - All business entities represented
  - N CompanyX applications covered
  - Connected to M databases
  - Identified NNN DB Schemas
  - Discovered NNNNN tables
  - Looked at NNNNNN data elements

Global IDs

- ★ **Read-only** access to live production tables
- ★ Ability to query database schema
- ★ Automatic data classification
- ★ Automatic data sampling and quality assessment

And more…

TRANSPARENCY

Looking at the real data

4.0 Profiling

Scenario

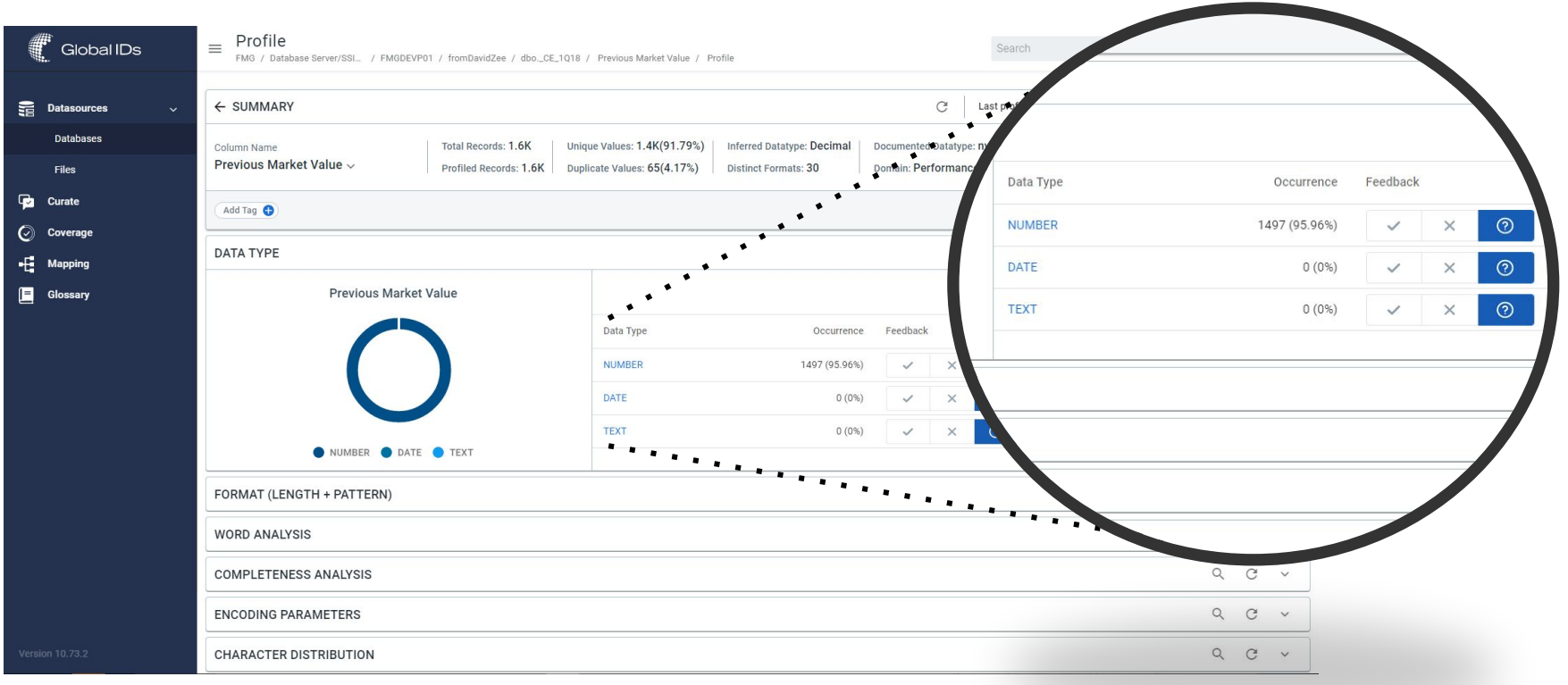
- ➔ No standard tools to profile data elements
- ➔ Custom scripting to QA data and address data issues
- ➔ DBEs are taking on operational roles
- ➔ No centralized program for data quality monitoring

Global IDs

- ★ Centralized view of all data elements
- ★ Various quality characteristics
- ★ Real-time information, updated as data is processed/received
- ★ Feedback system allowing users to correct systems' automated decision making

Example – Data Element Profiling

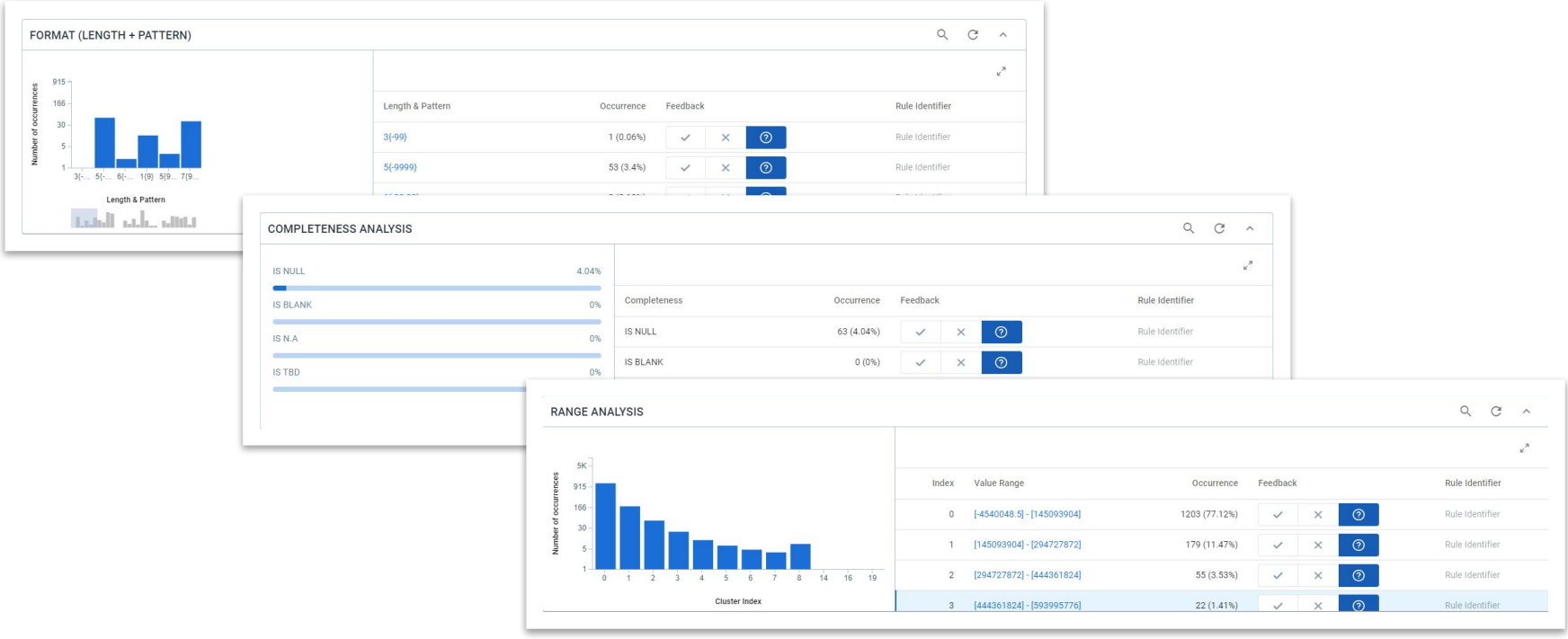
4.0 Profiling



TRANSPARENCY

Example – Data Element Profiling – contd.

4.0 Profiling



TRANSPARENCY

Automatic domain curation with user guidance

5.0 Classification

Scenario

➔ Currently no reliably documented mapping between table columns and business data domains exists

Global IDs

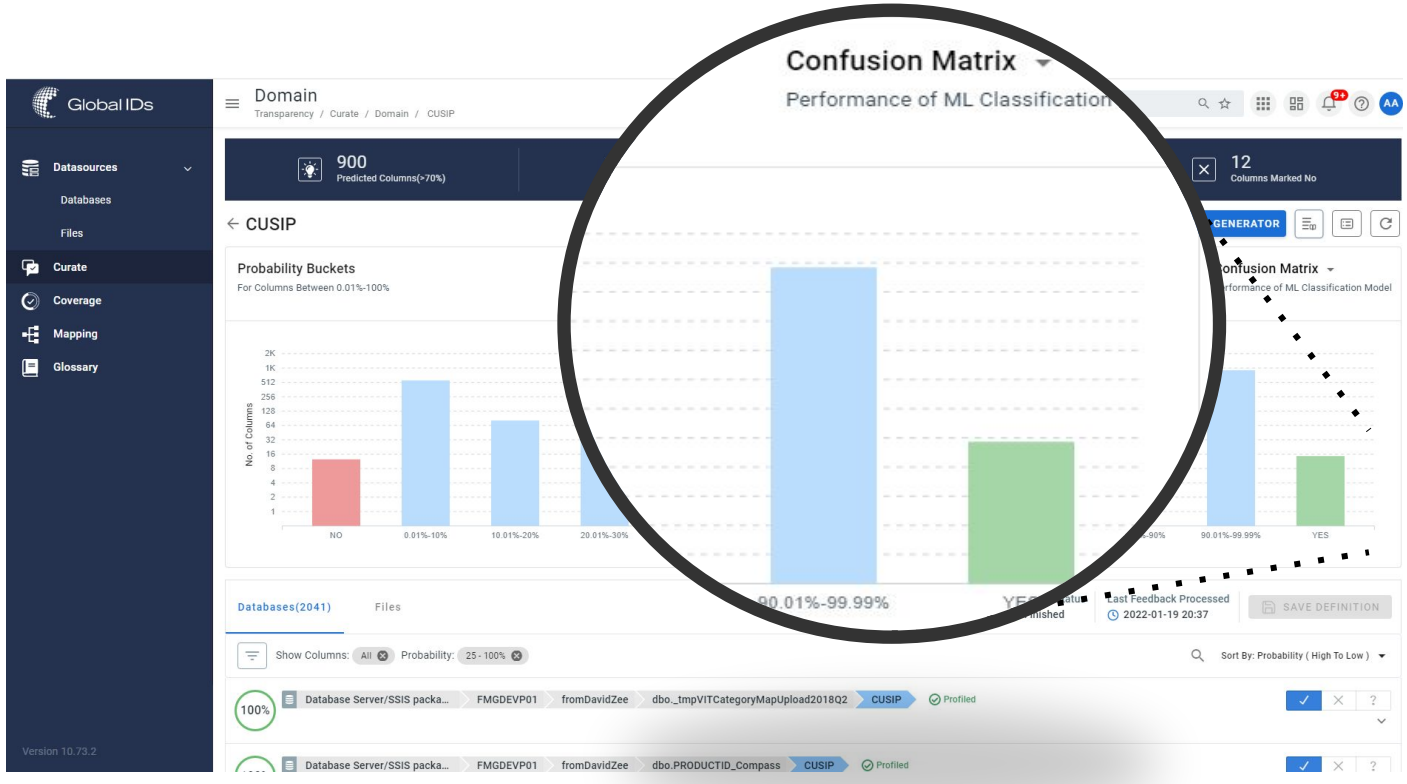
★ AI-based, automatically proposed mapping between database columns and semantic domains

★ User curation functionality supporting further reinforced learning

★ “Confidence level” mapping

★ Rapid feedback “Gaming” console

Example – Classification Engine



5.0 Classification

	PREDICTED TRUE	PREDICTED FALSE
POSITIVE	1157	188
NEGATIVE	890	146

Estimated True Positive
The number of columns marked YES plus the number of unmarked columns the system predicts will be marked YES.

Estimated False Positive
The number of columns the system may be erroneously predicting as YES considering the current accuracy level.

Estimated True Negative
The number of unmarked columns the system predicts will be marked NO.

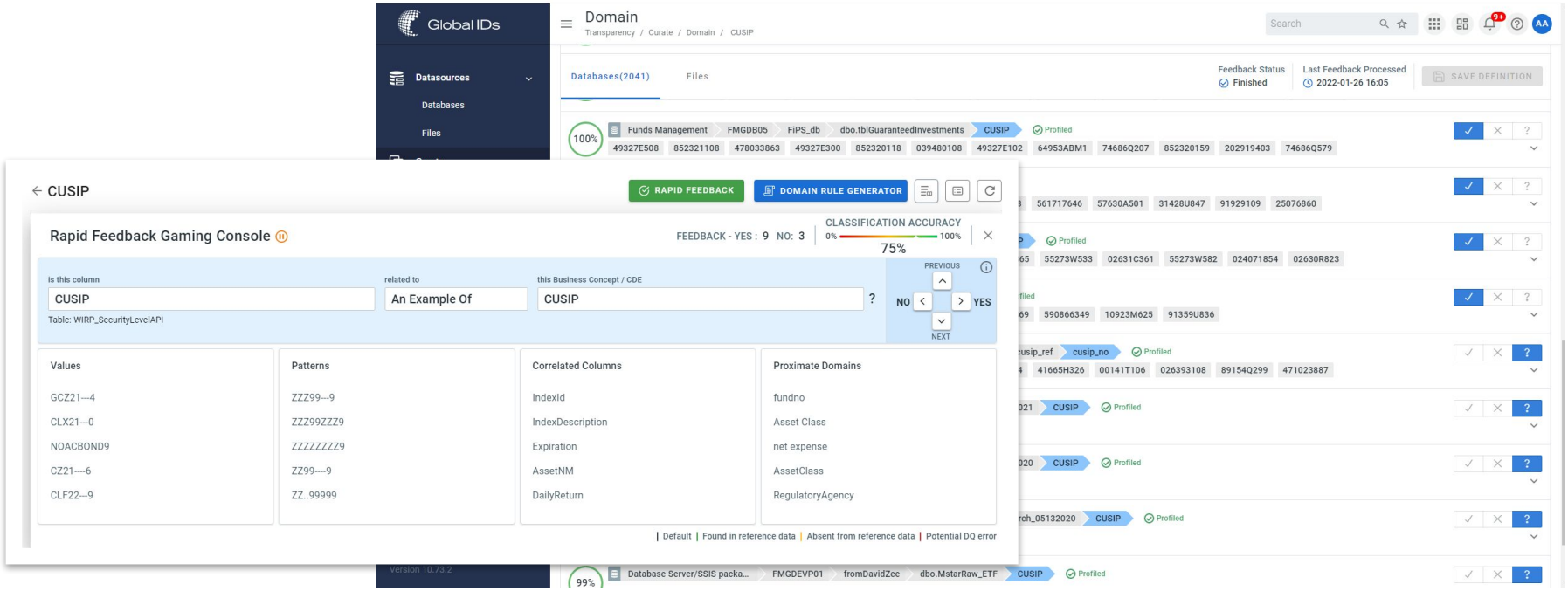
Estimated False Negative
The number of columns the system may be erroneously predicting as NO considering the current accuracy level.

80.01%-90% 90.01%-99.99%

TRANSPARENCY

Example – Classification Engine – user feedback

5.0 Classification



TRANSPARENCY

Transparency

We were able to consolidate metadata information in one place and map it to business domain terminology

- ➔ Business Ontology
- ➔ Semantic Objects
- ➔ Semantic Domains
- ➔ Database Columns
- ➔ Database Tables

Which business questions can we answer now?

- Where is my data?
- What database tables contain information about specific Business Object?
- What characteristics represent a specific Business Object?

Business concepts to database tables and fields mapping

6.0 Mapping

Scenario

➔ Currently no reliably documented mapping between table columns and business data domains exists

Global IDs

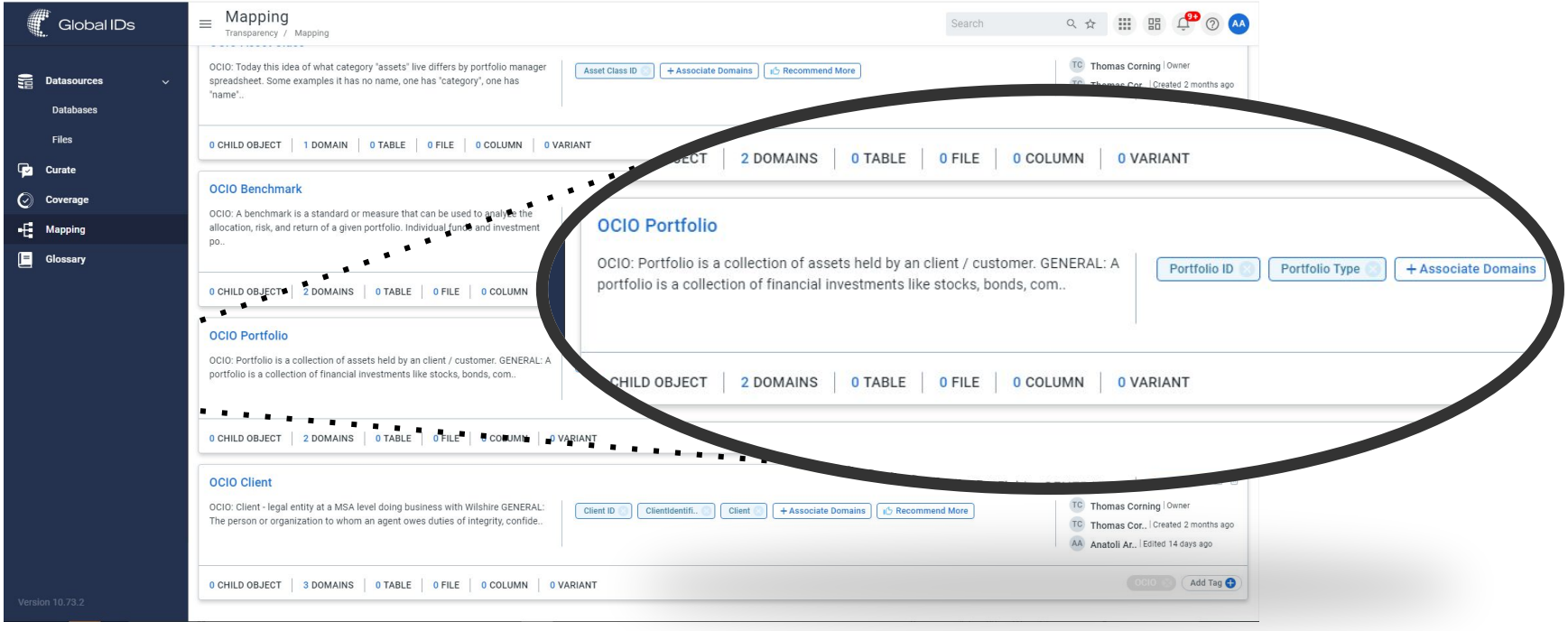
★ Semantic Objects (representing Business Terminology) are mapped to Semantic Domains (Representing Data Objects)

★ Semantic Domains are mapped to actual Data Columns of the Database Tables

E.g.:A business term **Portfolio** is represented by a semantic object **Portfolio**, which is represented by semantic domains of **Portfolio ID and Portfolio Name**, which are mapped to individual table columns

Example – Mapping Engine

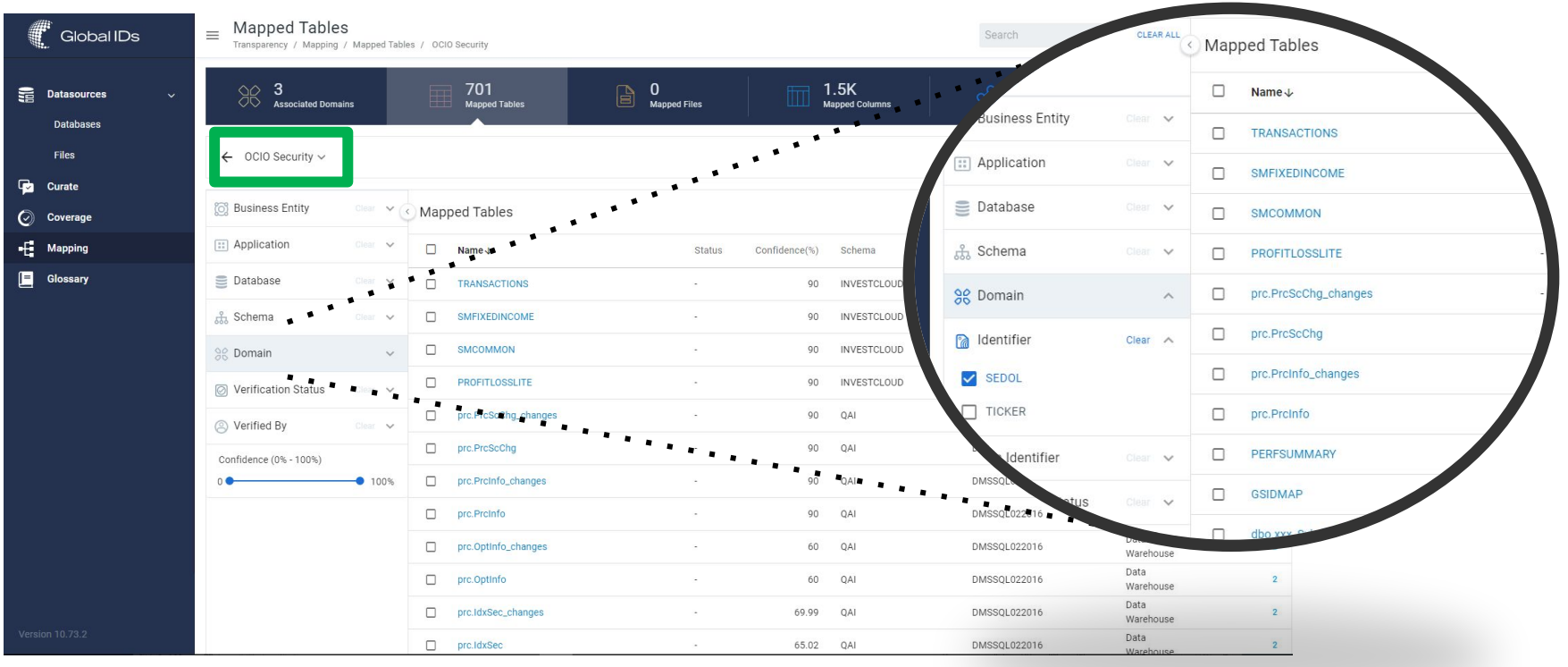
6.0 Mapping



TRANSPARENCY

Example – Mapped Tables (SOME Security)

6.0 Mapping



TRANSPARENCY

Analyzing and understanding data flows

7.0 Lineage

Scenario

➔ Currently no reliable or up-to-date data flow documentation exists

Global IDs

★ Automatically generated data flow and data lineage diagrams

★ Minimal directional user curation

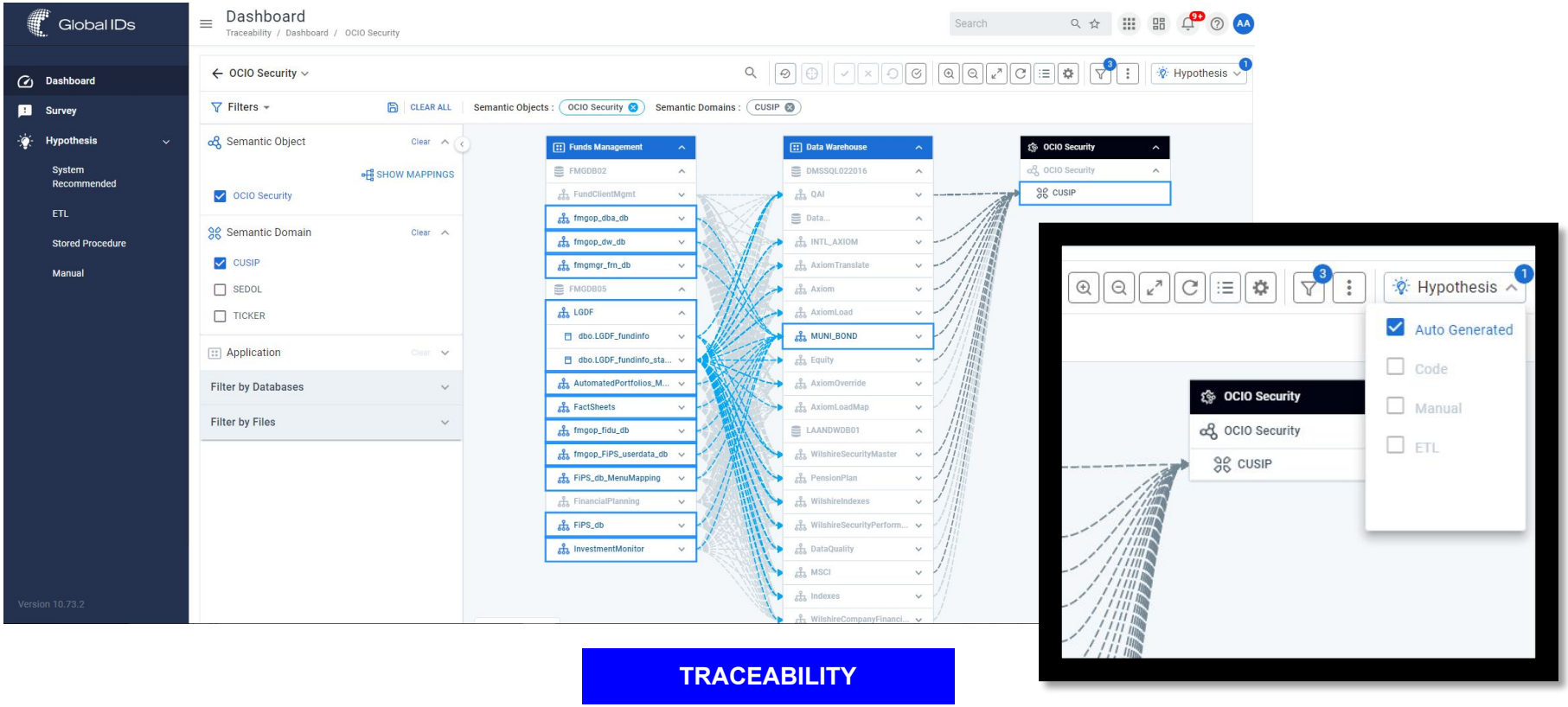
★ Visual representation of data flow with ability to highlight specific domains

★ Further ability to analyze code (stored procs), ETL configurations, etc.

TRACEABILITY

Example – Data Lineage

7.0 Lineage **8.0 Tracing**



TRACEABILITY

Traceability

We were able to connect applications with directional data flow information for specific semantic domains

- ➔ Data flow surveys within the context of an Application
- ➔ System data flow prediction based on the available metadata and user inputs
- ➔ Data traceability dashboards

Which business questions can we answer now?

- How does my data flow through the applications?
- Where does my data flow originate? Terminate?
- How is data relevant to a business term X routed through applications, databases?

Analyzing and ensuring data quality

9.0 Quality

Scenario

- ➔ Currently achieving data quality is a manual task
- ➔ Data quality process is not universal between groups and applications
- ➔ Operational data quality processes are in the hands of developers

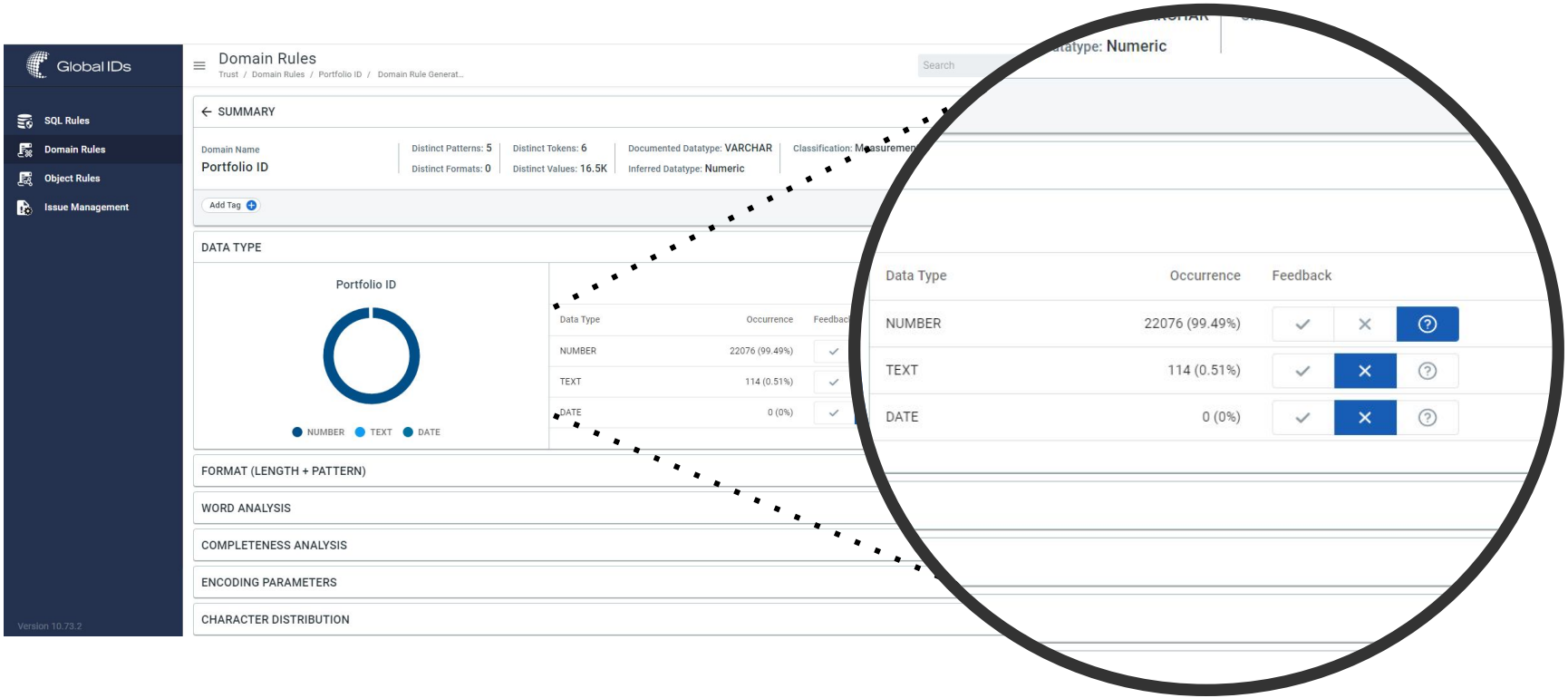
Global IDs

- ★ Comprehensive toolset allowing data quality monitoring:
  - ○ Domain rules
  - ○ SQL rules
  - ○ Semantic Object Rules
- ★ Issue management engine
  - ○ Create and assign issues directly from the data problems detected

TRUST

Example – Domain Rules

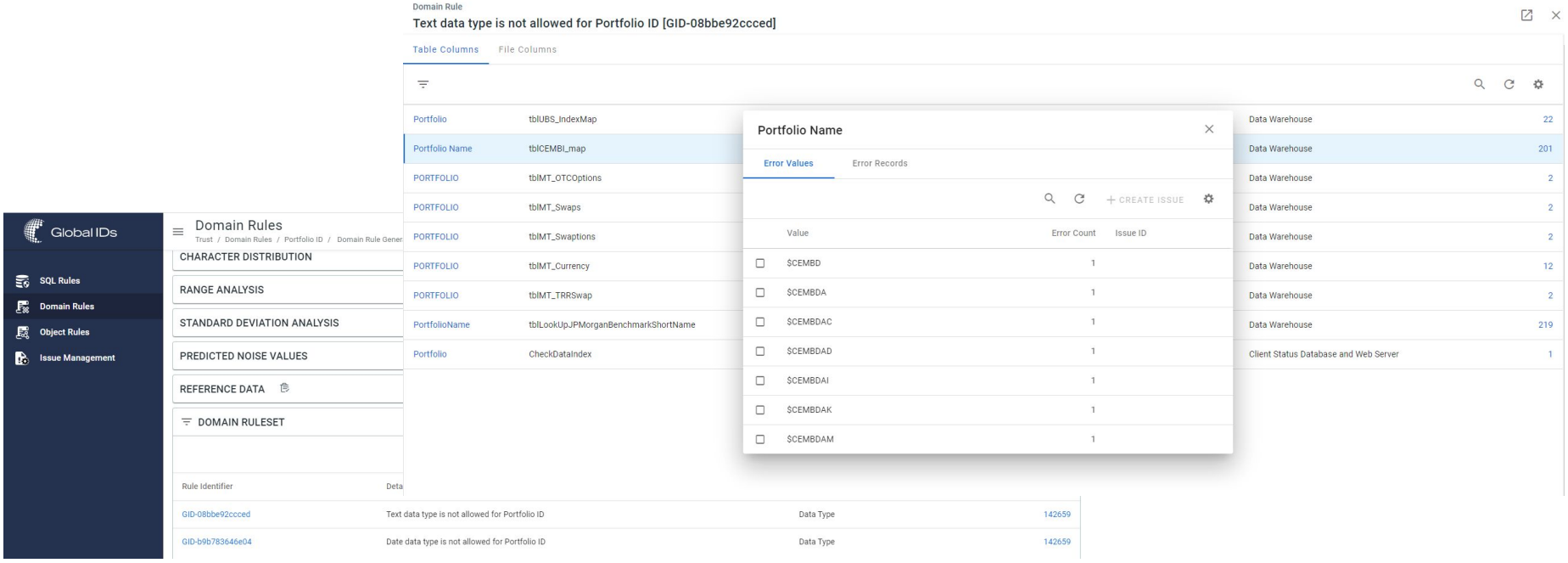
9.0 Quality



TRUST

Example – Domain Rules - Analysis

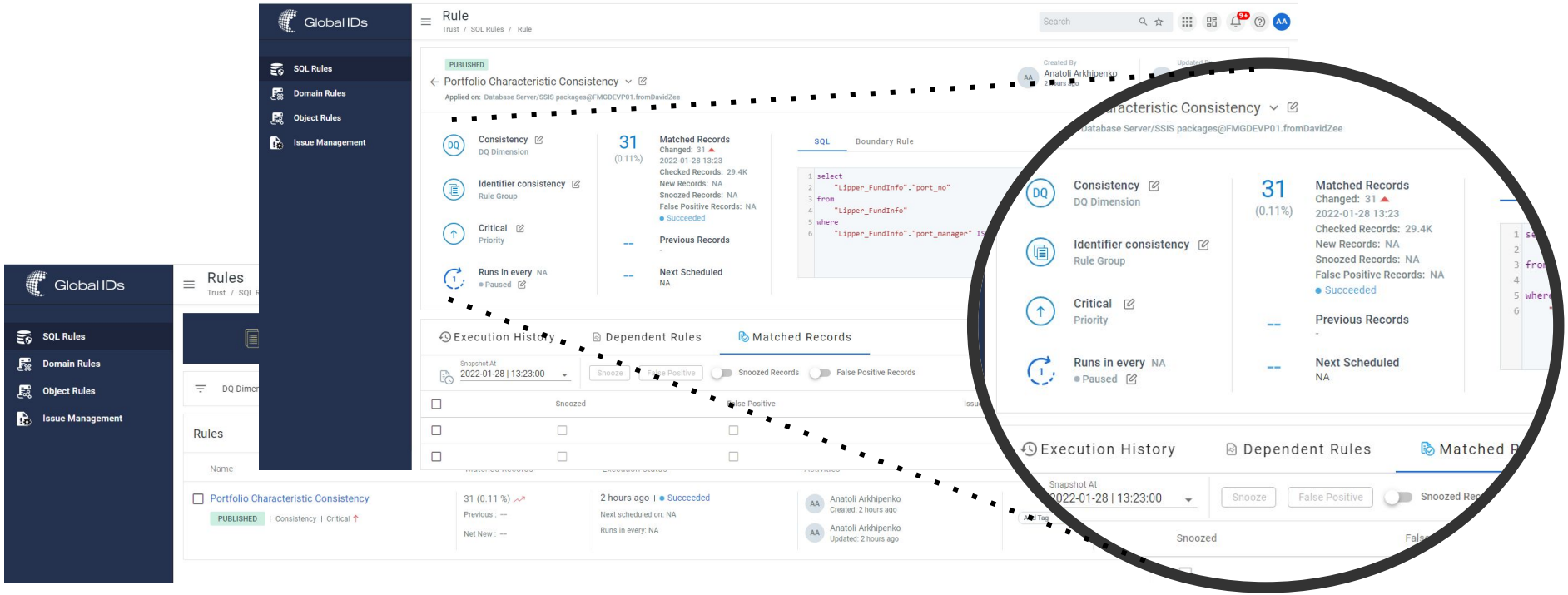
9.0 Quality



TRUST

Example – SQL Rules

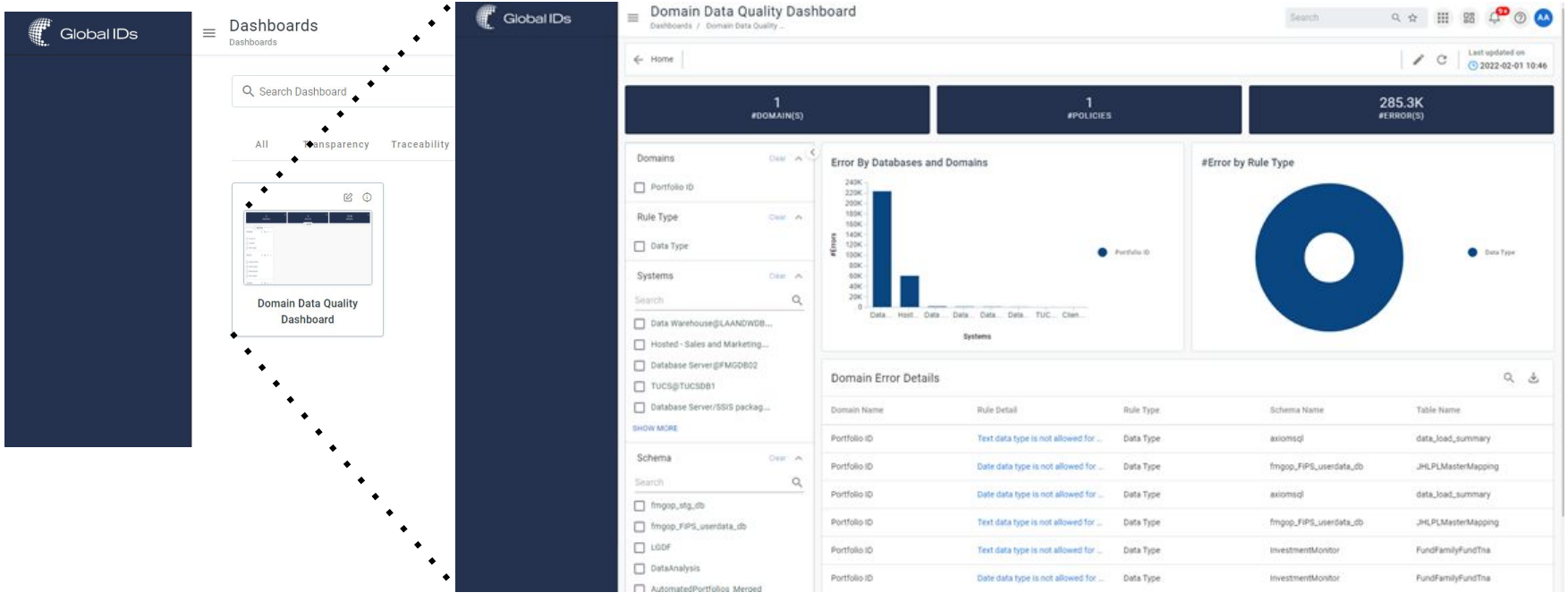
9.0 Quality



TRUST

Example – Data Quality Dashboards

10.0 Analytics



TRUST

Trust

We were able define specific validation rules for each semantic domain

- ➔ Various ways to validate data and metadata based on the real dataset
- ➔ Centrally monitor and react to data quality issues
- ➔ Spawn data issue resolution workflows

Which business questions can we answer now?

- Can I trust my data?
- Is data coming from vendors accurate and complete?
- Where do I have data quality issues?
- Who is working on a data quality issue X?