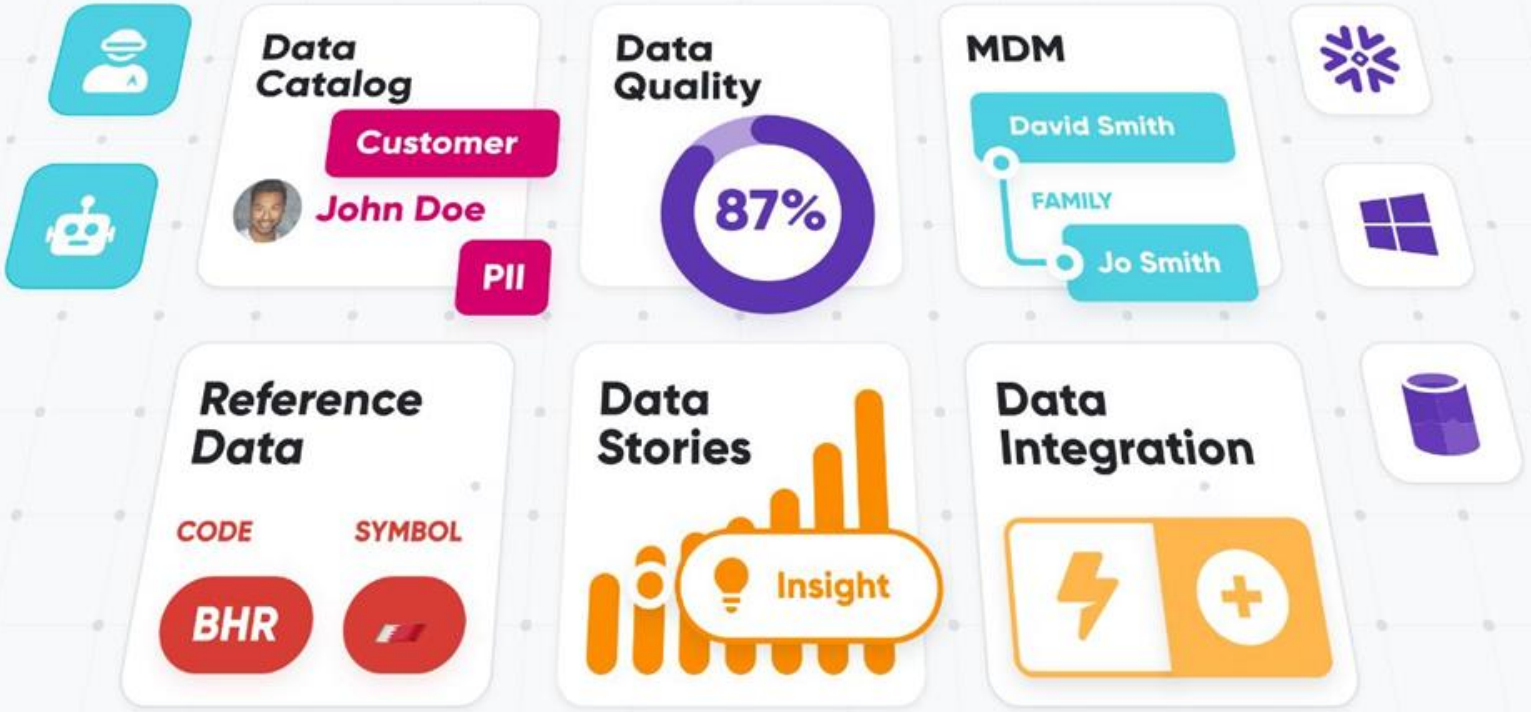
ataccama

Data Quality Fabric

Make your enterprise data instantly available to humans and machines as high quality data products, with governance and compliance ensured automatically.



TODAY‘S

Agenda

1. **Company Introduction**
2. **From Integrated to Self-driving**
3. **Platform overview**
4. **Data Catalog**

ATACCAMA

Company introduction

Ataccama Fast Facts

Offices

Toronto, London, Sydney, Prague
New York, Valencia, Paris, Sofia

350+ Global Customers

In 23 distinct countries
+ 15 states

60,000+ DQA Downloads

Gartner Recognition

Both DQ and MDM Gartner Magic Quadrant

Top positions in Peer Insights

Company Highlights

New Offices

Sydney, Paris

Over 200% growth

In 3 Years

New Partners

Technology, Platform, SI

Professional Services

Capacity doubled

Cloud Offering

Managed Services, PaaS
Marketplaces

HCA
Hospital Corporation of America

gsk

BlueCross
BlueShield

Daiichi-Sankyo

American
Airlines

CANADIAN
TIRE

Lifeboats

GE Capital

First Data

T · · Mobile ·

BMO

O2
A Telefónica company

vodafone

TD Bank

novozymes

ARGO
GROUP

Allianz

AVIVA

SOCIETE
GENERALE

avast

Raiffeisen
BANK

HSBC

Winnipeg

Otsuka

Walmart

Auchan

beazley

FIFTH THIRD BANK

tpl
toronto
public library

Celgene

SELECTED CUSTOMERS

Financial Services

- Allianz
- Argo Group
- Aviva
- Bank of Montreal
- Beazley
- Fifth Third Bank
- First Data Corporation
- GE Capital Corporation
- HSBC
- Industrial Alliance
- KBC Group
- Raiffeisenbank Group
- RSA
- Scotiabank
- Societe Generale
- TD Bank

SELECTED CUSTOMERS

Pharma
Healthcare
Public

- BCBS Association
- BCBS Michigan
- BCBS Minnesota
- Celgene Corporation
- City of Winnipeg
- Daiichi Sankyo
- Edwards Lifesciences
- Glaxo Smith Kline
- Hospital Corporation of America
- Otsuka Pharmaceutical
- MedAmerica Inc.
- Memorial Health System
- Novozymes

SELECTED CUSTOMERS

Telco
Retail
Other

- American Airlines
- Auchan
- AVAST
- Canadian Tire
- CD Cargo
- Darden Restaurants
- GAF
- Helmerich and Payne

- Husky Energy
- Intersnack
- Telefonica O2
- RNLI
- SOCAN
- T-Mobile
- Toronto Public Library
- Vodafone

Ataccama partner ecosystem

Technology, OEM and cloud partners

OEM distribution
Software compatibility & joined value proposition
Customized Cloud-based solutions
Data enrichment services

Consulting Partners Value-Added Resellers

Integration & consulting services
Reselling Ataccama SW, custom-tailored solutions
Alliance Partner Program members or
Opportunity-Based Global and region-specific

TIBCO
collibra
ADASTRA
UNISYS
TATA CONSULTANCY SERVICES
LOQATE
cloudera
Ask Bigger Questions
TERADATA
Raising Intelligence
wipro
Capgemini
KPMG
Deloitte.
aws
Azure
Google Cloud
accenture
VIQTOR DAVIS
DATA CRAFTSMANSHIP
Cognizant

Ataccama Service Offering

Ataccama is more than software. Our expanding professional services team ensures your success and maximum value out of your Ataccama platform investment. Our services cover the entire solution lifecycle.

Enablement

- Training
- Installation
- Knowledge base
- Starter pack

Strategic

- Architecture
- Governance
- Advanced configuration
- Trouble shooting

Rollout

- Configuration
- Integration
- Consulting support

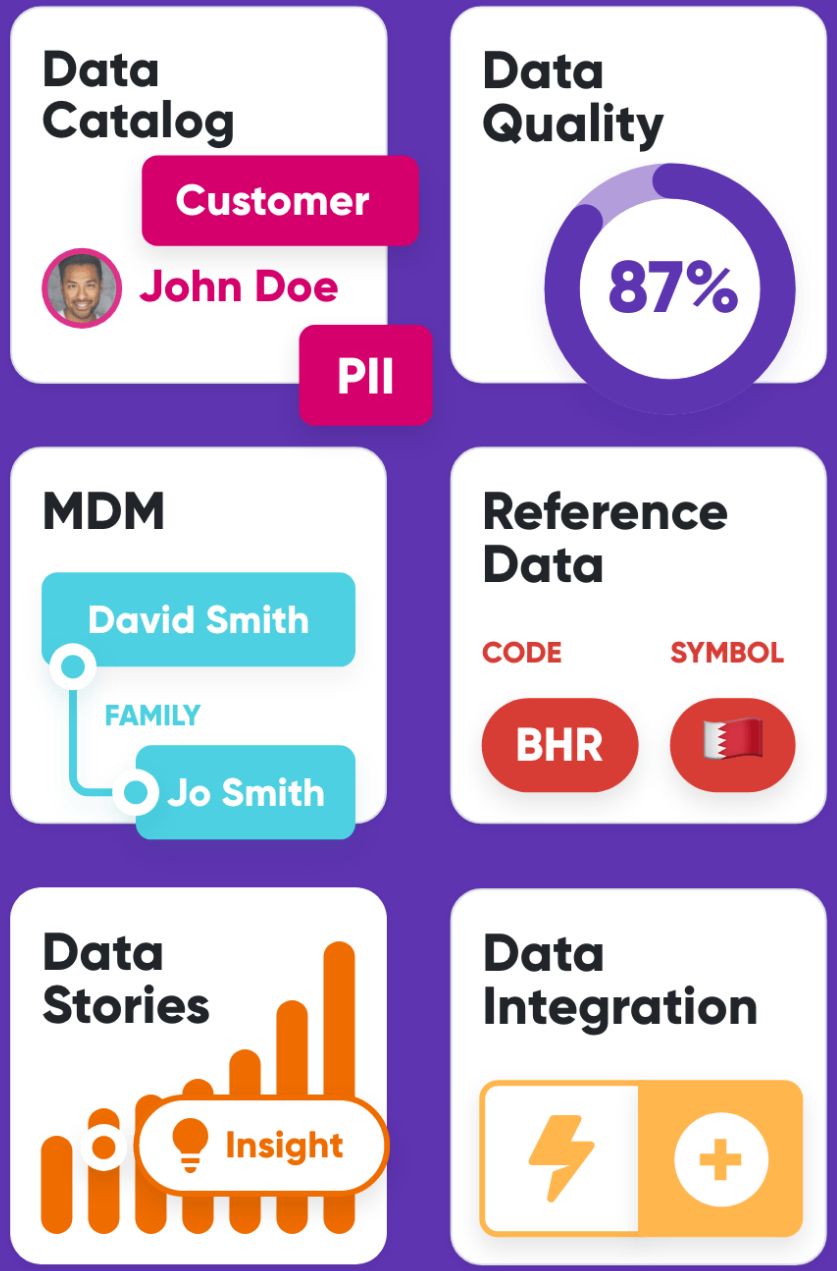
Improvement

- Tuning
- Optimization
- Health-check
- Upgrade

Support

- Helpdesk
- Operations
- Online documentation
- Community

Enterprise Data Quality Fabric



Make high quality data instantly available to

Data engineers

Data stewards

Value creators

Machines & algorithms

With governance and compliance ensured automatically.

Selected Use Cases

Data Governance

Data Profiling

Data Catalog

Metadata Repositories

Data Quality

DQ Monitoring & Reporting

Data Visualization

DQ Firewall

Data Cleansing & Enrichment

Manual Issue Resolution

Data Integration

Multi-domain MDM

Reference Data Management

Multi-Domain Mastering

Master Domain Monitoring

Master Data Authoring

Householding & Aggregation

Consent Management

EPIM

Product Information Management

Media Asset Management

Platform Differentiators

Unified Platform with Complementary Modules

Active metadata usage – Extendibility – Embedded Governance

Self-Service Solution

AI & Machine Learning – Self-Service – Collaboration – UX/UI

Enterprise-Proven Usage

Scalability – Performance – High Availability – Auditing – Identity Management

Robust Data Processing Engine

Any Data / Any Domain – Integration – Performance – Scalability

Security

ENCRYPTION

Transit -> TLS 1.2

Rest -> native to Azure/AWS RDBMS

AUTHENTICATION

AD via SAML (incl. Azure/AWS)

AUTHORIZATION

Keycloak (part of Ataccama ONE platform) manages mapping from AD groups/roles/claims

LOGGING

Full audit log

Deployment options

Platform as a Service

Focus on your data initiatives while we take care of the infrastructure, monitoring, maintenance, and security.

Marketplaces

Deploy the Platform from a cloud marketplace of your choice with a just few clicks.

On Prem & Hybrid

Deploy on your own infrastructure, private & public clouds and their combinations.

Built for cloud data management, deploy natively in the cloud, big data platforms, or hybrid environments

Platform as a Service

Focus on your data initiatives and bring value to your organization while we take care of the infrastructure, maintenance, and security

Marketplaces

Deploy the platform from a cloud marketplace of your choice.

Private Cloud

Deploy our containerized solution into your private cloud with Docker & Kubernetes.

Supporting all major cloud platforms

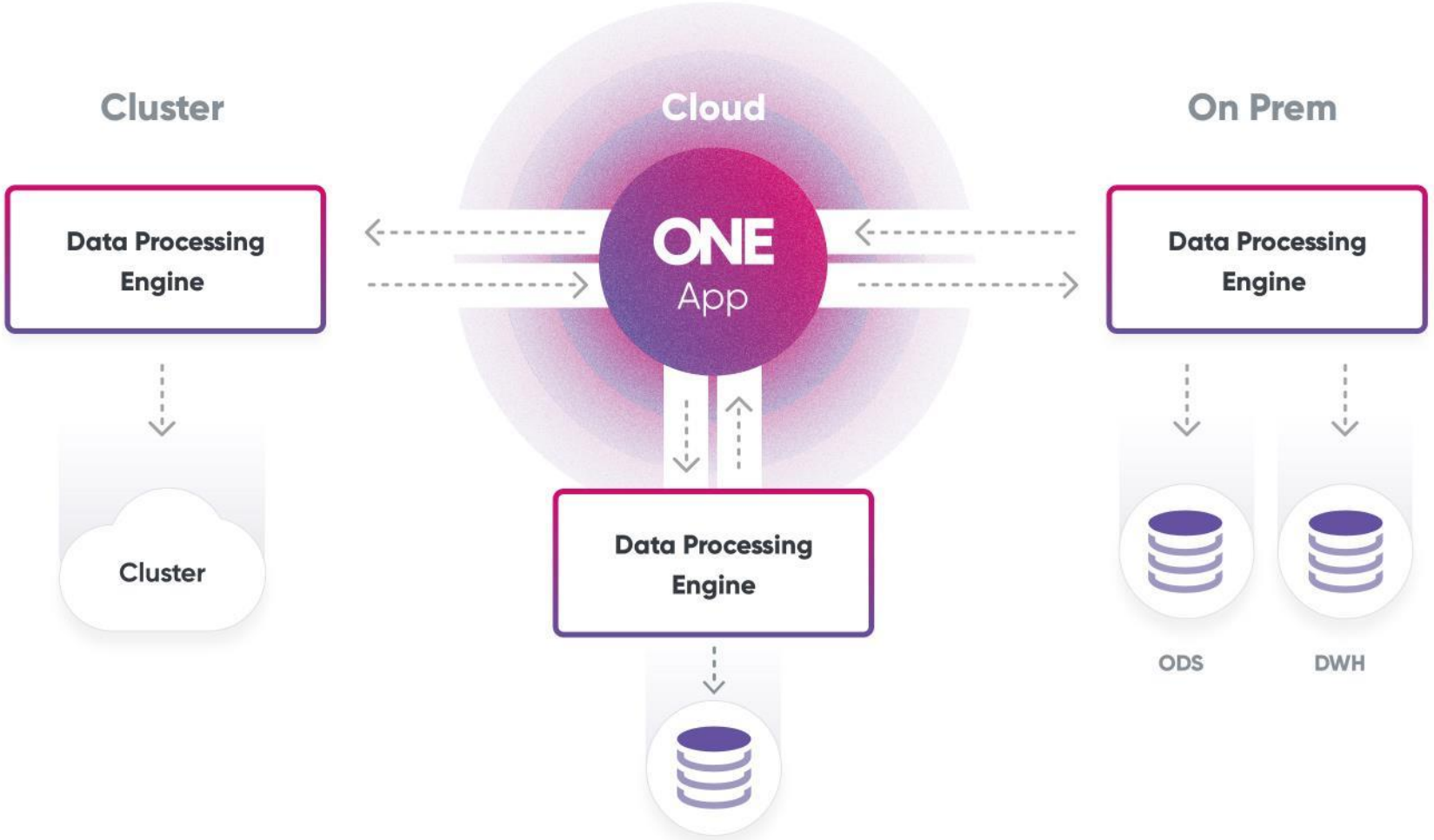
aws

Microsoft Azure

Google Cloud

Hybrid deployment

ATACCAMA ONE



Metadata Management

Data Catalog

Data Catalog Drivers

Lack of visibility into the data landscape

Multiple sources
& scattered data

Different technologies

Too much time spent on getting data

Finding the right source
and owner

Getting access or exports

Data preparation

Lack of trust in data

Can I trust this source?

Is this data of high
enough quality?

What kind of data is
inside?

Ataccama’s Solution

Capture

Automatically discover Metadata

Import metadata

Manually create metadata

Manage

Detect and manage business domains

All data assets in a knowledge catalog

Curate and update

Provide

Search for assets

Access underlying data

Use it directly or provide via APIs

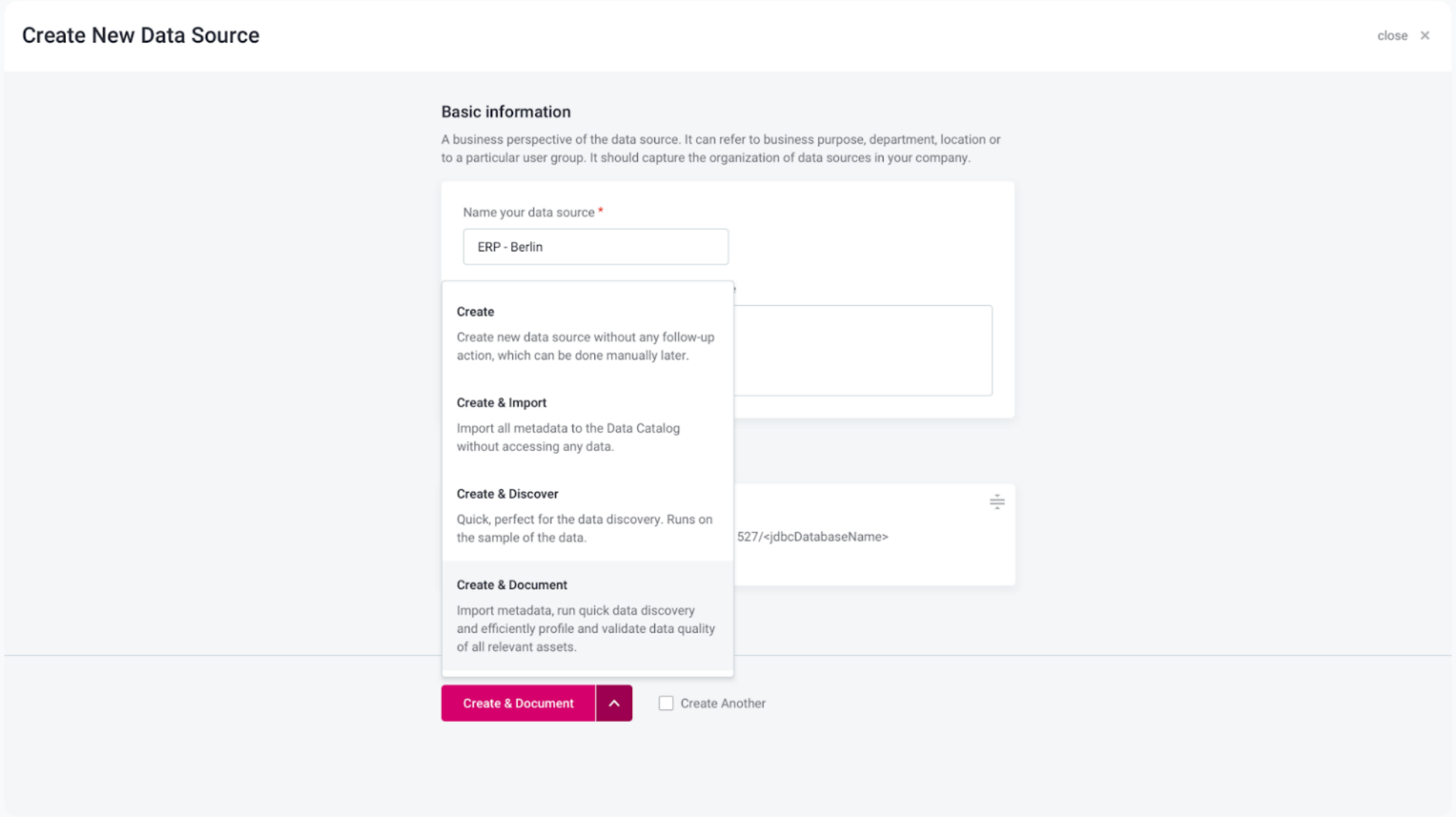
Ataccama Differentiators

Self—updating view of your data landscape

Search and filter by any metadata

Built-in data quality

Connect all your data sources


Create New Data Source
close
Basic information
A business perspective of the data source. It can refer to business purpose, department, location or to a particular user group. It should capture the organization of data sources in your company.
Name your data source *
ERP - Berlin
Create
Create new data source without any follow-up action, which can be done manually later.
Create & Import
Import all metadata to the Data Catalog without accessing any data.
Create & Discover
Quick, perfect for the data discovery. Runs on the sample of the data.
Create & Document
Import metadata, run quick data discovery and efficiently profile and validate data quality of all relevant assets.
527/<jdbcDatabaseName>
Create & Document
Create Another

All key information about your data
in one place and always up to date

KNOWLEDGE CATALOG

OVERVIEW

DATA CATALOG

Data Assets

Business Systems

Data Sources

Data Marts

Infrastructure

Rules

Components

BUSINESS GLOSSARY

BUSINESS PROCESSES

PROJECTS

Home > Knowledge Catalog > Sources

ERP - Berlin

Document

Overview | Catalog Items | Data Quality | Lineage | Connection Details

2

Overall Quality	Amount of data	Connections	Schemas	Catalog items
75%	6,7 TB	2	19	1 007
↙ 3%	↗ 1,2 TB		↗ 3	↗ 77

Documented Catalog Items

Imported	1 001	100%
Sampled	757	75%
Profiled	234	23%
Data Quality	212	21%

Detected domains

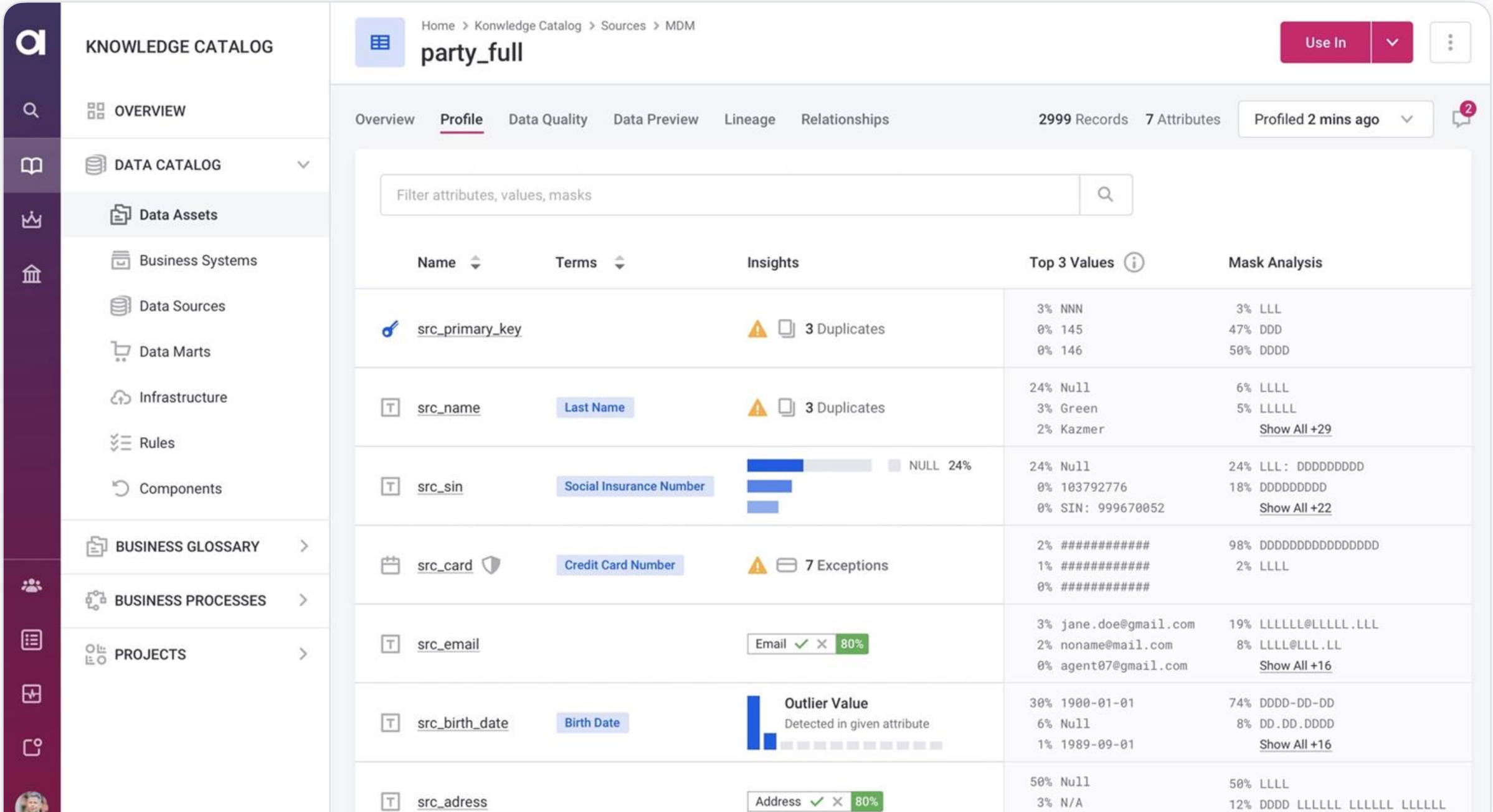
Parent of

Domain	Table Occurences
Personal Information	765

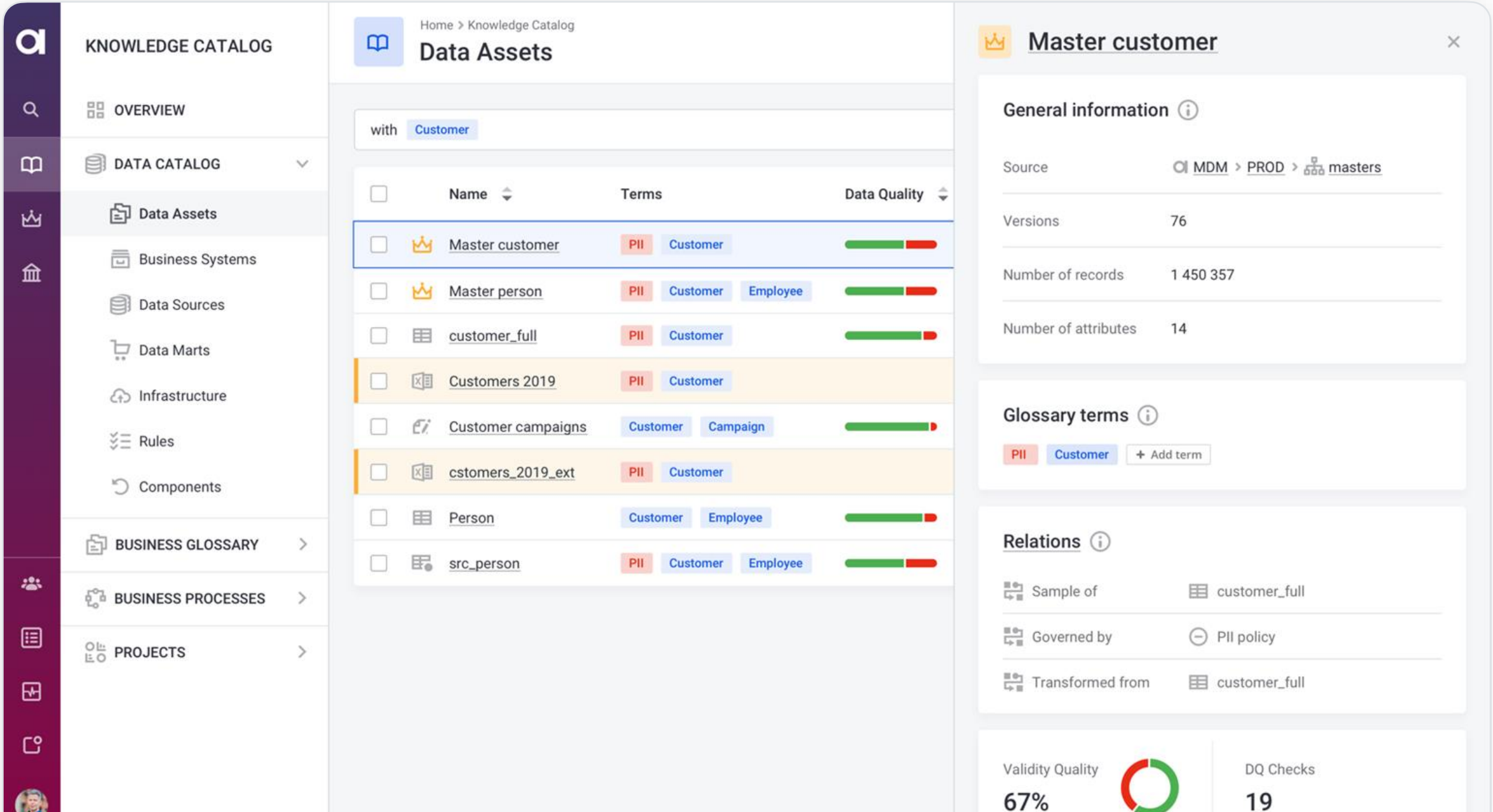
General Information

Enterprise resource planning (ERP) is the integrated management of main business processes, often in real time and mediated by software and technology.

Status	ACTIVE
Type	Primary
Business category	ERP, Billing, Planning
Department	DT3026 - Governance / Application
Business owner	Sammy Lawson
Technical owner	Sidnee Gye


KNOWLEDGE CATALOG
OVERVIEW
DATA CATALOG
Data Assets
Business Systems
Data Sources
Data Marts
Infrastructure
Rules
Components
BUSINESS GLOSSARY
BUSINESS PROCESSES
PROJECTS
Home > Konwledge Catalog > Sources > MDM
party_full
Use In
Overview
Profile
Data Quality
Data Preview
Lineage
Relationships
2999 Records
7 Attributes
Profiled 2 mins ago
2
Filter attributes, values, masks
Name
Terms
Insights
Top 3 Values
Mask Analysis
src_primary_key
3 Duplicates
3% NNN
0% 145
0% 146
3% LLL
47% DDD
50% DDDD
src_name
Last Name
3 Duplicates
24% Null
3% Green
2% Kazmer
6% LLLL
5% LLLLL
Show All +29
src_sin
Social Insurance Number
NULL 24%
24% Null
0% 103792776
0% SIN: 999670052
24% LLL: DDDDDDDDD
18% DDDDDDDDD
Show All +22
src_card
Credit Card Number
7 Exceptions
2% ############
1% ############
0% ############
98% DDDDDDDDDDDDDDDD
2% LLLL
src_email
Email
80%
3% jane.doe@gmail.com
2% noname@mail.com
0% agent07@gmail.com
19% LLLLLL@LLLLL.LLL
8% LLLL@LLL.LL
Show All +16
src_birth_date
Birth Date
Outlier Value
Detected in given attribute
30% 1900-01-01
6% Null
1% 1989-09-01
74% DDDD-DD-DD
8% DD.DD.DDDD
Show All +16
src_adress
Address
80%
50% Null
3% N/A
50% LLLL
12% DDDD LLLLLL LLLLLL LLLLLL

Find the right data fast with search and filtering


KNOWLEDGE CATALOG
OVERVIEW
DATA CATALOG
Data Assets
Business Systems
Data Sources
Data Marts
Infrastructure
Rules
Components
BUSINESS GLOSSARY
BUSINESS PROCESSES
PROJECTS
Home > Knowledge Catalog
Data Assets
with Customer
Name
Terms
Data Quality
Master customer PII Customer
Master person PII Customer Employee
customer_full PII Customer
Customers 2019 PII Customer
Customer campaigns Customer Campaign
cstomers_2019_ext PII Customer
Person Customer Employee
src_person PII Customer Employee
Master customer
General information
Source MDM > PROD > masters
Versions 76
Number of records 1 450 357
Number of attributes 14
Glossary terms
PII Customer + Add term
Relations
Sample of customer_full
Governed by PII policy
Transformed from customer_full
Validity Quality
67%
DQ Checks
19

Use our AI and make it work better for you

Suggestions for Personal information

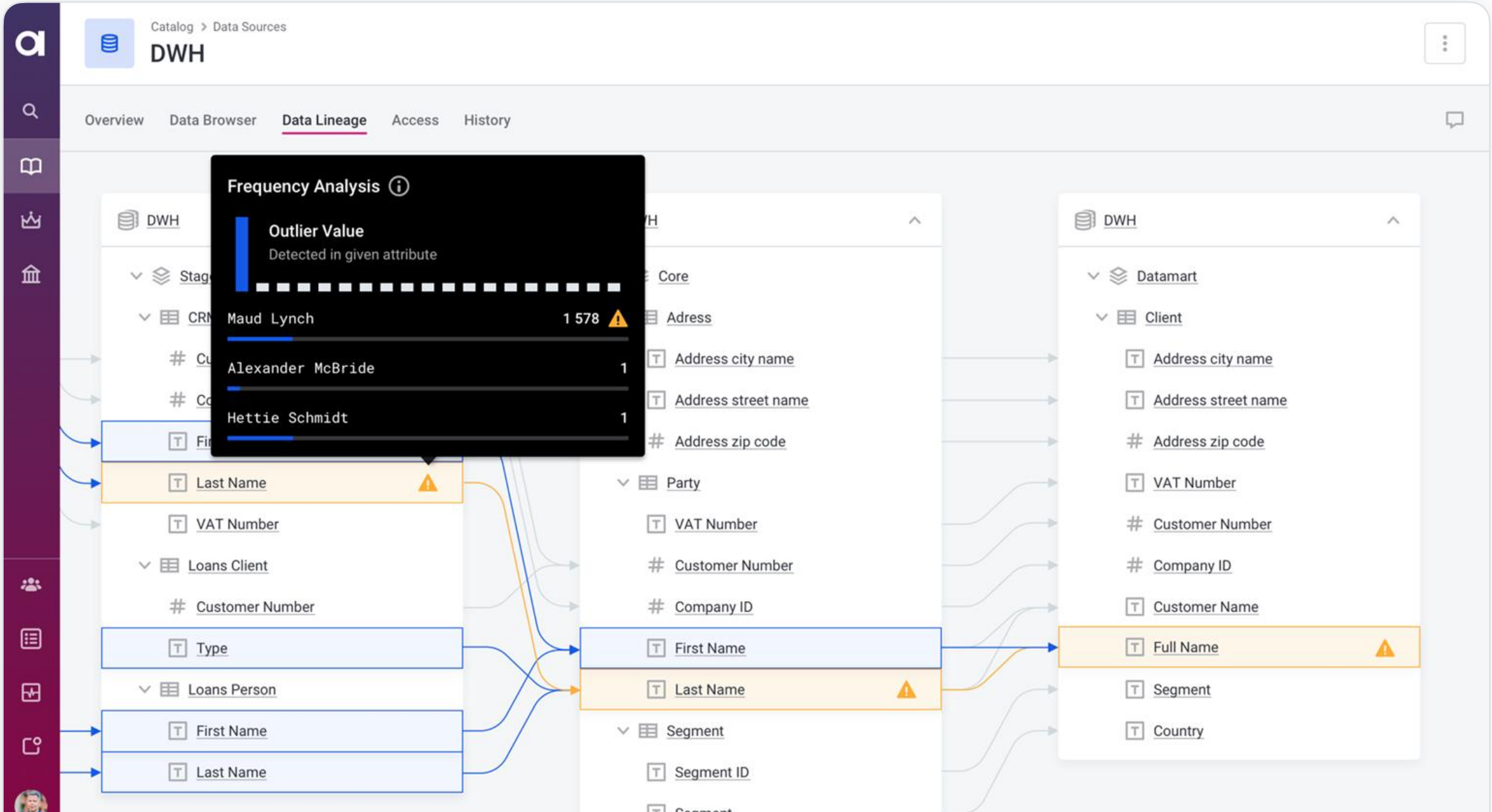
Filter by suggestion, name or a term

Name	Terms	Confidence	Source	
displayname	Name	79%	EMEA > Retail > Customers	Undo
birth_date	Date of birth	77%	EMEA > Retail > Customers	Accept Reject
phone_number	Phone	77%	EMEA > Retail > Customers	Accept Reject
zip	ZIP Code	72%	Alpha > GIM > Person	Accept Reject
email	E-mail address	71%	EMEA > Retail > Customers	Accept Reject
bd	Date of birth	68%	Alpha > GIM > Person	Accept Reject
contact	Phone	66%	Alpha > GIM > Person	Accept Reject
zip	ZIP Code	62%	SalesForce marketing cloud > Google Analytics > Web	Accept Reject
contact_information	E-mail address	61%	Alpha > GIM > Person	Accept Reject

Describe your changes

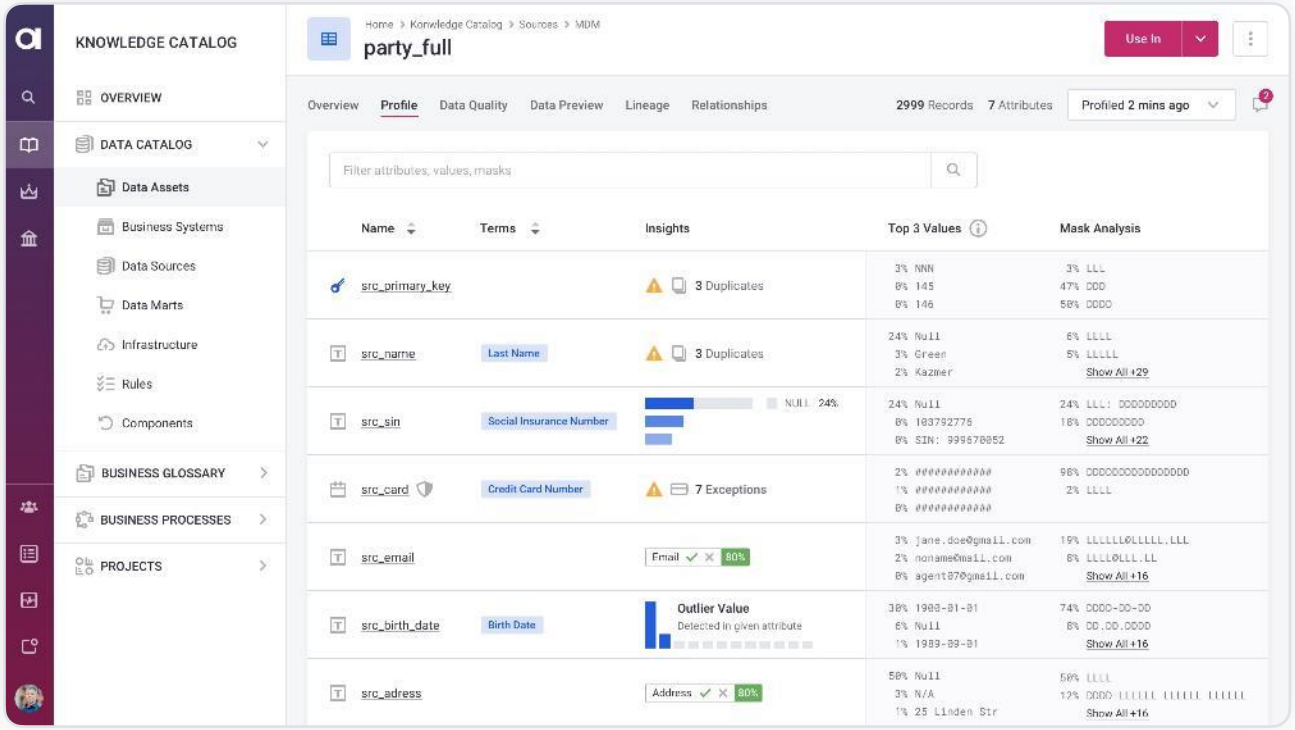
Save & Submit for approval

Discard all changes


Catalog > Data Sources
DWH
Overview
Data Browser
Data Lineage
Access
History
Frequency Analysis
Outlier Value
Detected in given attribute
Maud Lynch 1 578
Alexander McBride 1
Hettie Schmidt 1
DWH
Last Name
VAT Number
Loans Client
Customer Number
Type
Loans Person
First Name
Last Name
Core
Adress
Address city name
Address street name
Address zip code
Party
VAT Number
Customer Number
Company ID
First Name
Last Name
Segment
Segment ID
DWH
Datamart
Client
Address city name
Address street name
Address zip code
VAT Number
Customer Number
Company ID
Customer Name
Full Name
Segment
Country

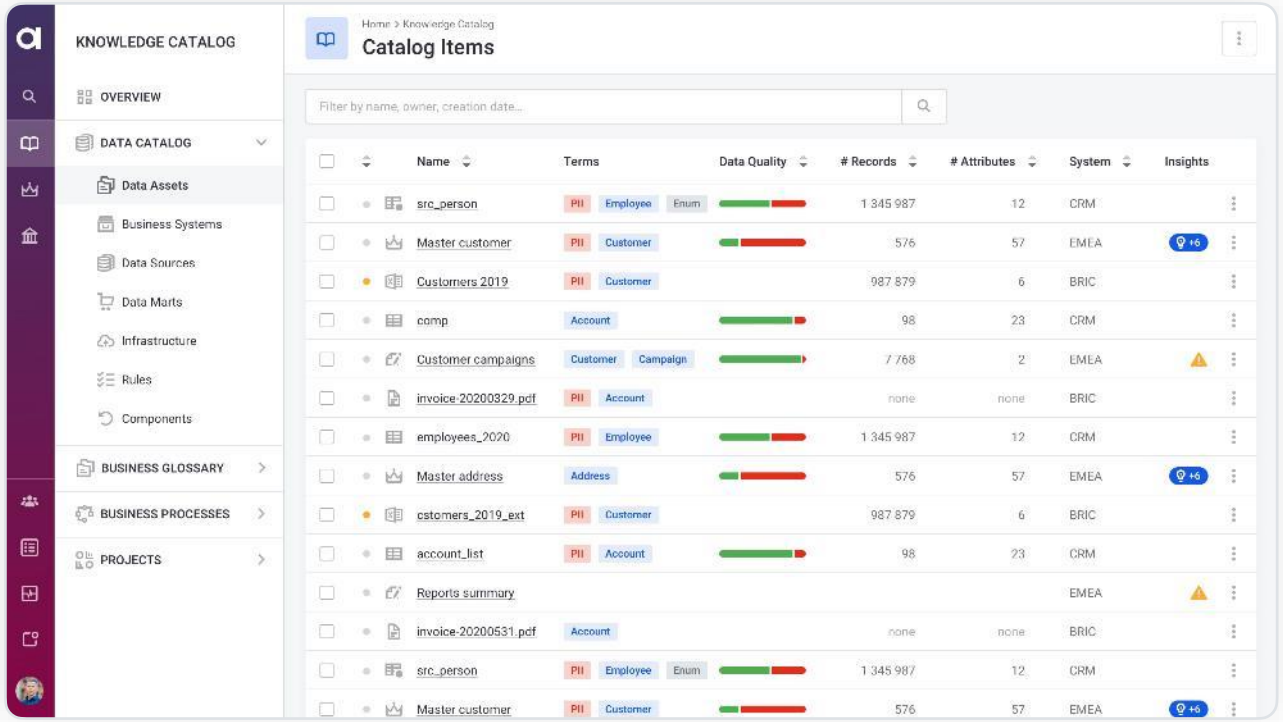
Smart, flexible,
and automated business glossary

Data profiling



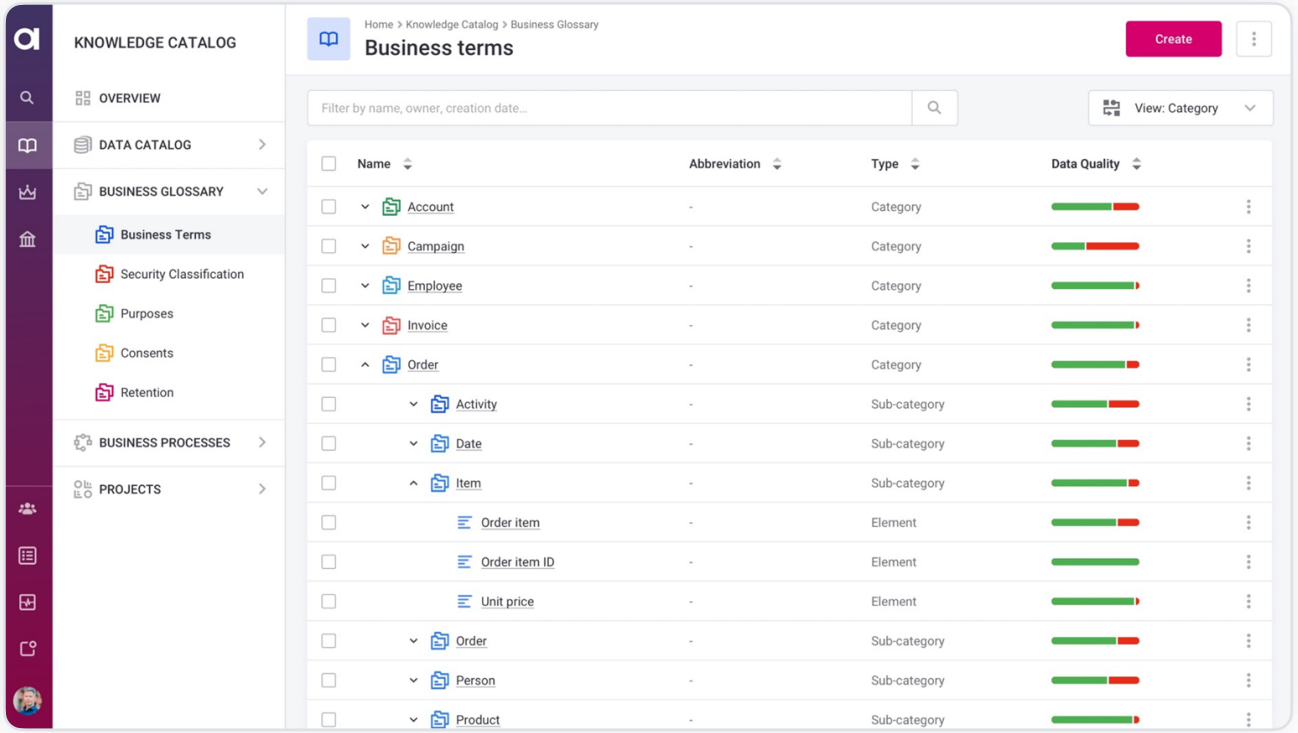
- Automated out of the box profiling
- Data domain discovery
  - AI-driven classification
  - SME-driven process
- Automated detection of outliers and exceptions
- Sharing of profile reports

Knowledge catalog



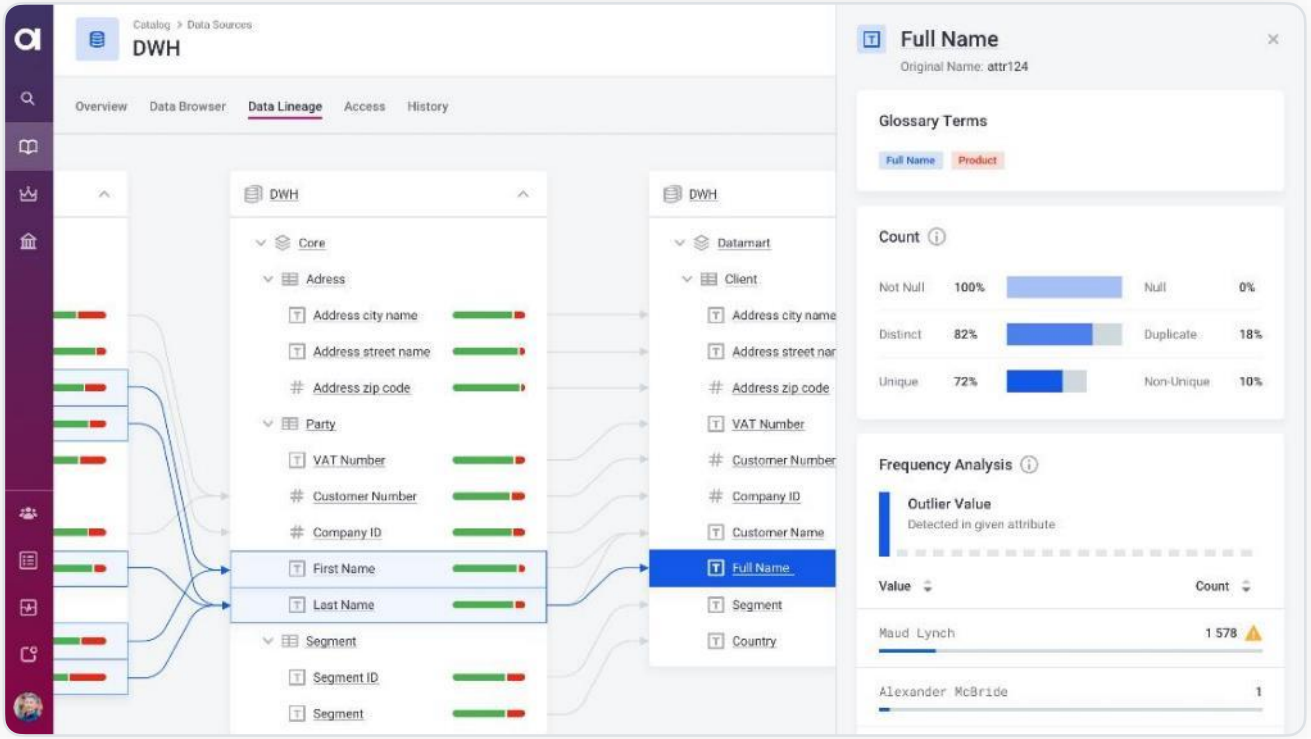
- Locate, capture and store metadata
  - Automatically discovered metadata
  - Imported 3rd party metadata
  - Manually created metadata
- Capture data relations and lineage
- Automatically calculate DQ metrics
- Share metadata assets across the organization

Business glossary



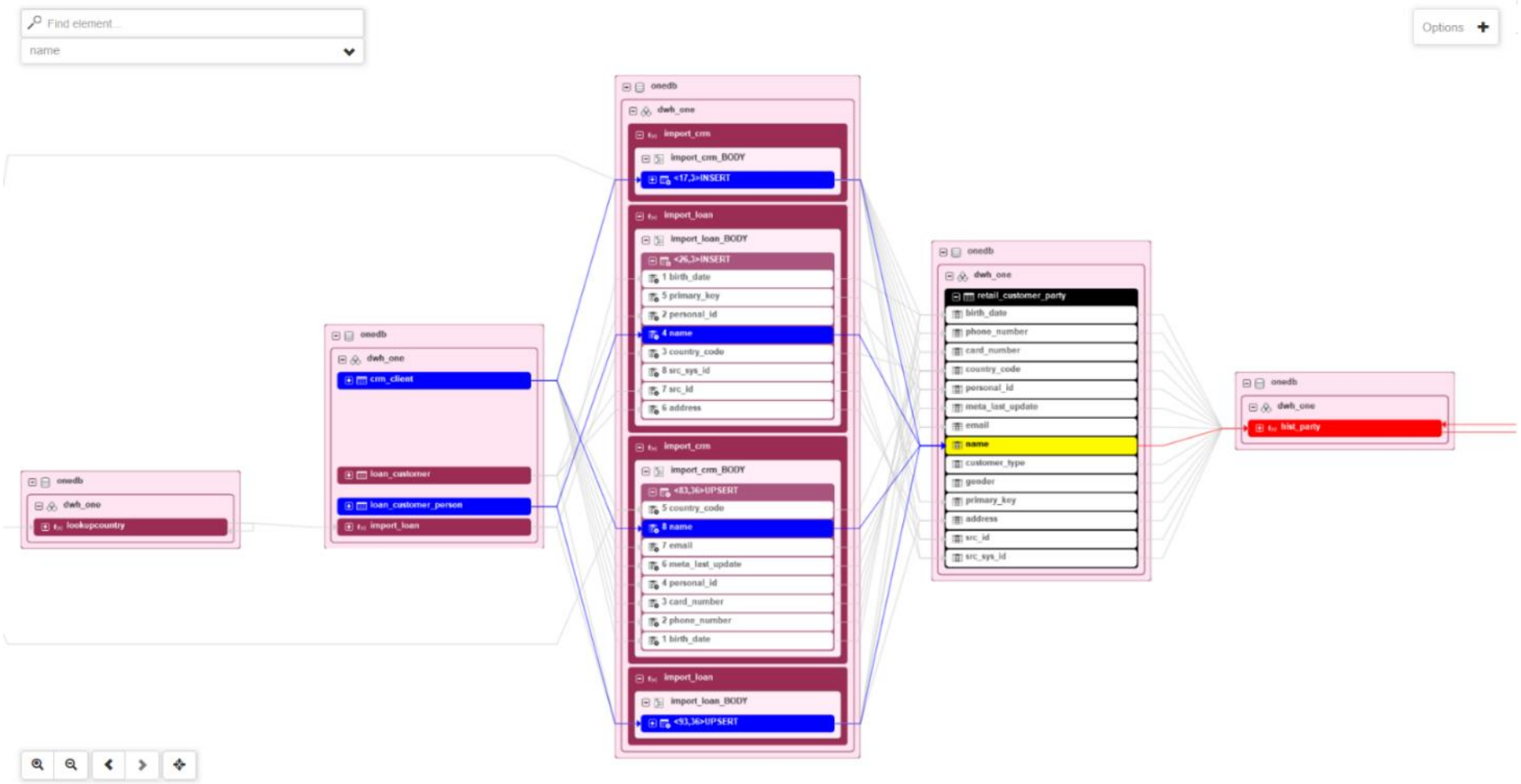
- Central component of Ataccama ONE managing all business terms
- Relationships among business terms are possible and their hierarchy is available
- Occurrence statistics across all connected data sources
- Advanced conditions or rules as part of the business term definition
- Link with relevant data assets, links and stakeholders

Data lineage

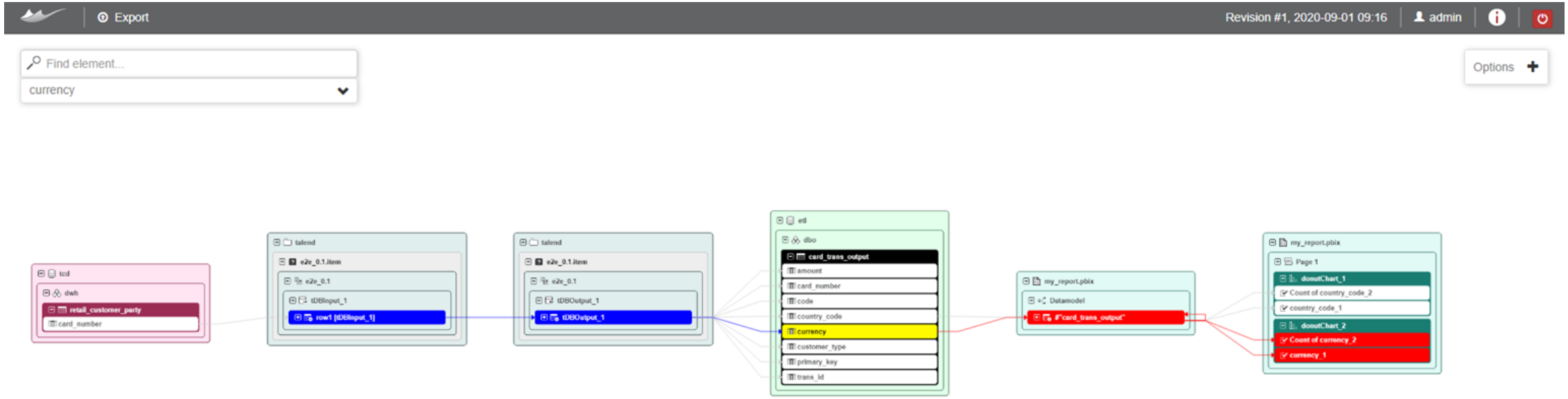


- Overlays with business terms, data quality indicators, and anomalies
- Track a data problem to its source
- Focus on a specific business term and find the right data set
- Import lineage from external tools
- Full technical lineage also available

Full technical detail

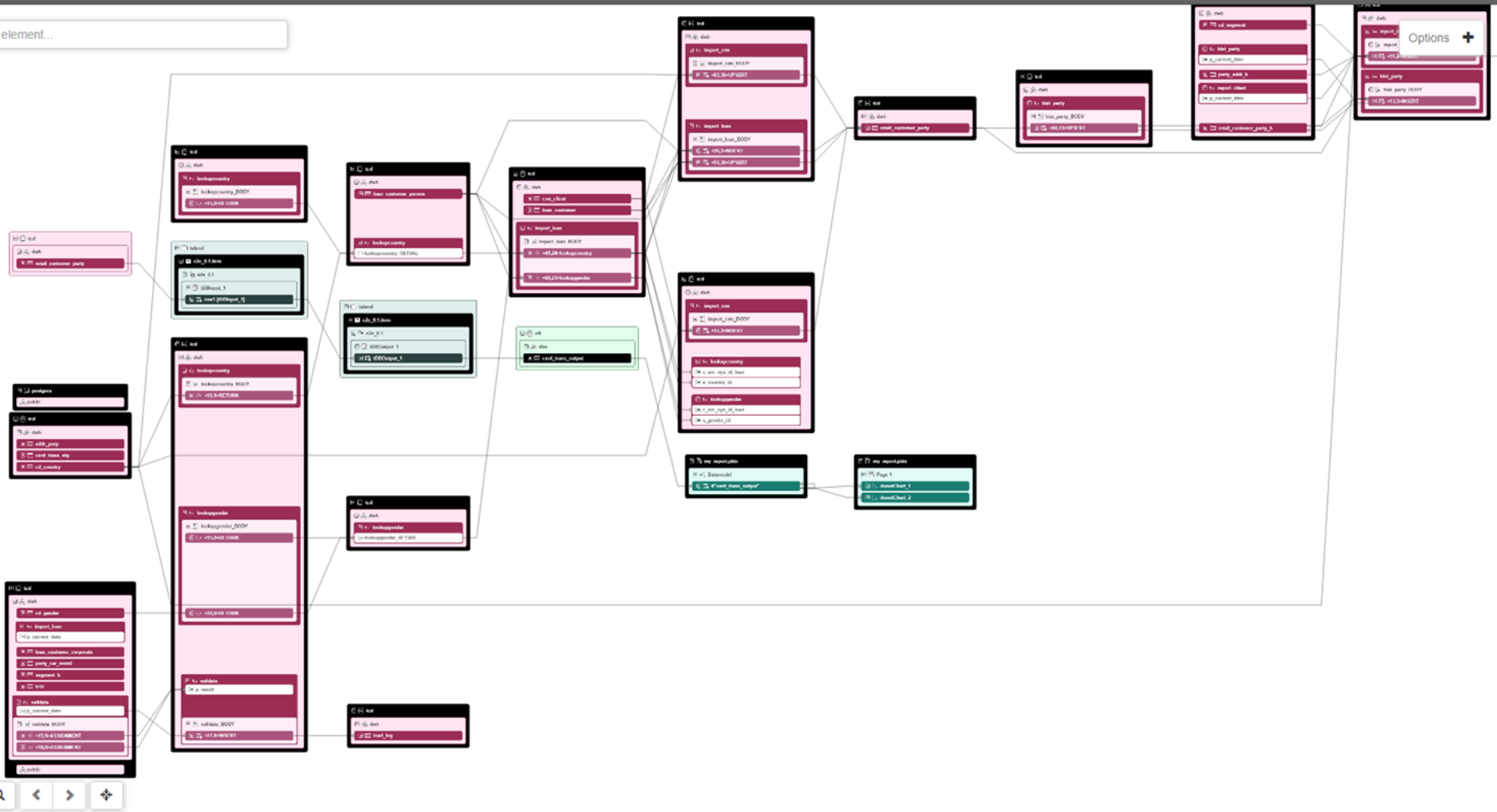


Full technical detail

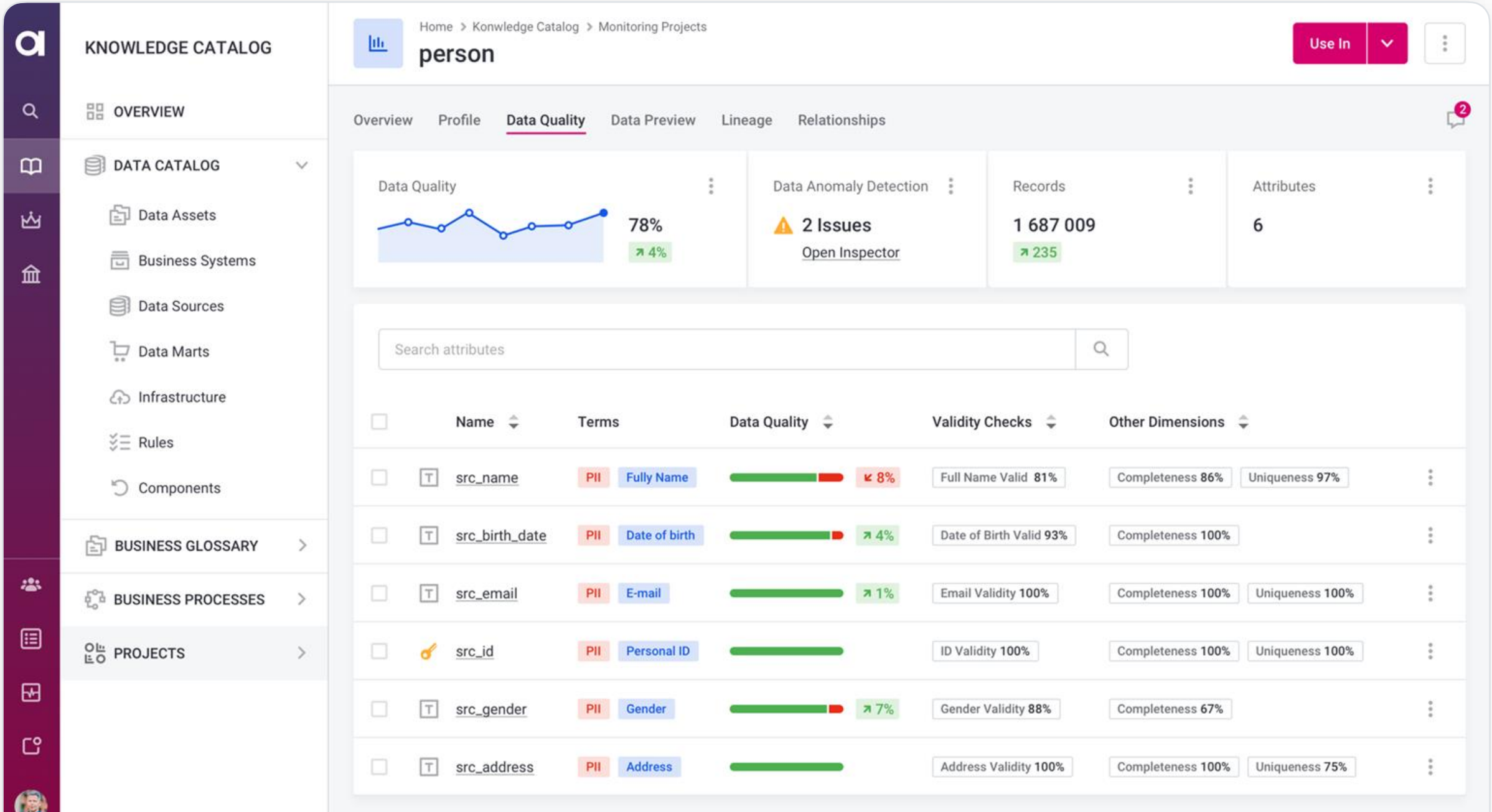


Find element...

Options +



Built-in, always up to date data quality for any entity


KNOWLEDGE CATALOG
OVERVIEW
DATA CATALOG
Data Assets
Business Systems
Data Sources
Data Marts
Infrastructure
Rules
Components
BUSINESS GLOSSARY
BUSINESS PROCESSES
PROJECTS
Home > Konwledge Catalog > Monitoring Projects
person
Use In
Overview
Profile
Data Quality
Data Preview
Lineage
Relationships
2
Data Quality
78%
4%
Data Anomaly Detection
2 Issues
Open Inspector
Records
1 687 009
235
Attributes
6
Search attributes
Name
Terms
Data Quality
Validity Checks
Other Dimensions
src_name
PII
Fully Name
8%
Full Name Valid 81%
Completeness 86%
Uniqueness 97%
src_birth_date
PII
Date of birth
4%
Date of Birth Valid 93%
Completeness 100%
src_email
PII
E-mail
1%
Email Validity 100%
Completeness 100%
Uniqueness 100%
src_id
PII
Personal ID
ID Validity 100%
Completeness 100%
Uniqueness 100%
src_gender
PII
Gender
7%
Gender Validity 88%
Completeness 67%
src_address
PII
Address
Address Validity 100%
Completeness 100%
Uniqueness 75%

KNOWLEDGE CATALOG

OVERVIEW

DATA CATALOG

BUSINESS GLOSSARY

Business Terms

Security Classification

Purposes

Consents

Retention

BUSINESS PROCESSES

PROJECTS

Home > Konwledge Catalog > Glossary

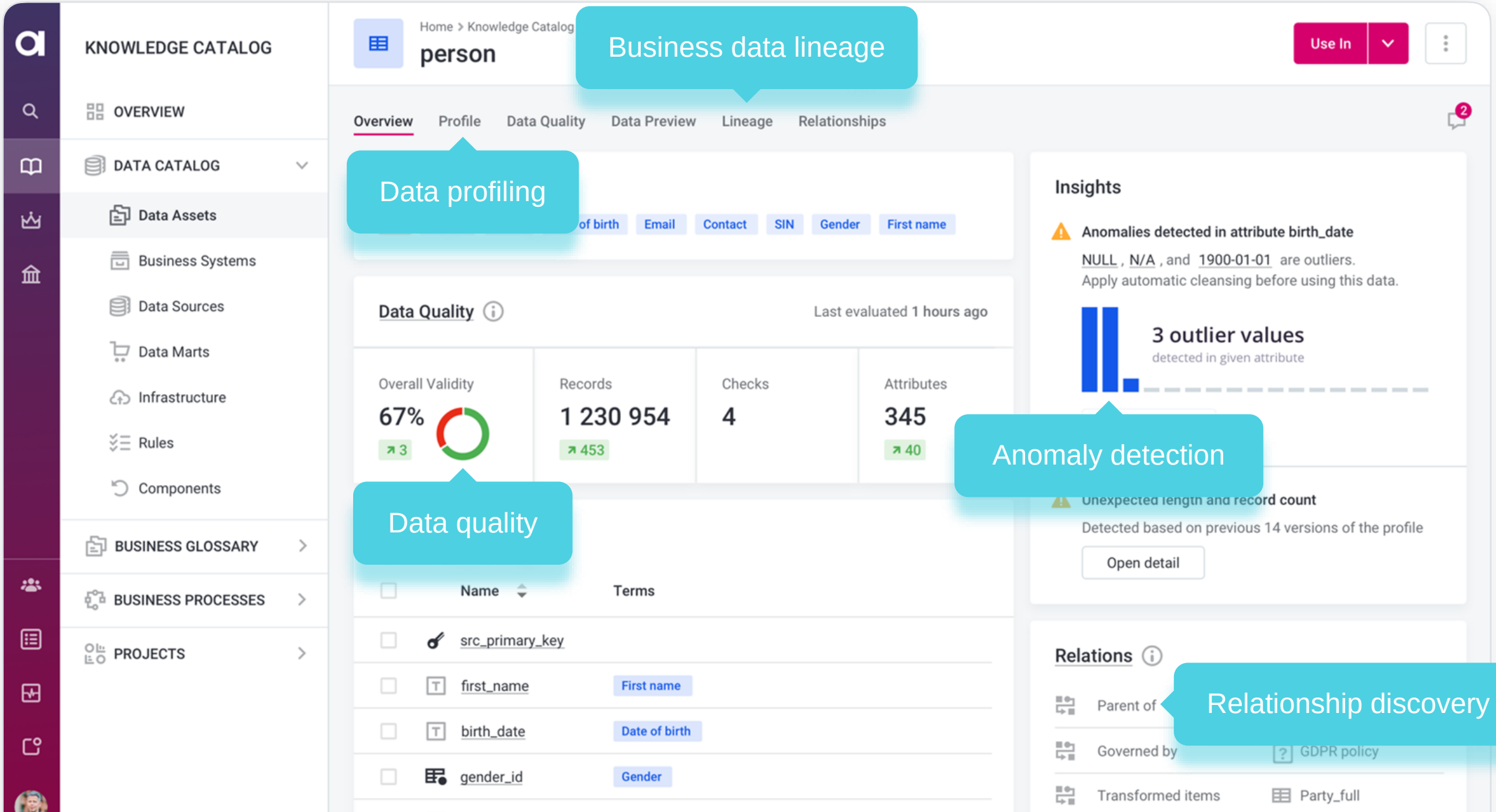
E-mail

Overview | Relationships | Occurance | Data Quality | Settings

Aggregated Validity	Records	Data Sources	Catalog Items	Attributes
75% ↙ 8%	1 687 009 ↗ 235	29	67 ↗ 5	73 ↗ 5

Search attributes

Attribute	Term	Overall Validaty		Other Dimensions	
email	E-mail	↙ 12%	⚠ Email Validity 65%	⚠ Email Validity 65%	Uniqueness 86%
E-mail	E-mail	↗ 4%	Email Validity 81%	Completness 84%	Uniqueness 100%
e-mail	E-mail	↗ 1%	Email Validity 100%	Completness 100%	Uniqueness 100%
attr14	E-mail		Email Validity 100%	Completness 100%	Uniqueness 100%
src_mail	E-mail	↗ 7%	Email Validity 80%	Completness 90%	Uniqueness 67%
attr14	E-mail		Email Validity 100%	Completness 100%	Uniqueness 100%
src_mail	E-mail	↗ 3%	Email Validity 81%	Completness 90%	Uniqueness 73%
contact	E-mail	↙ 8%	Email Validity 76%	Completness 98%	Uniqueness 100%
contact	E-mail	↙ 2%	Email Validity 96%	Completness 95%	Uniqueness 87%


KNOWLEDGE CATALOG
OVERVIEW
DATA CATALOG
Data Assets
Business Systems
Data Sources
Data Marts
Infrastructure
Rules
Components
BUSINESS GLOSSARY
BUSINESS PROCESSES
PROJECTS
Home > Knowledge Catalog
person
Use In
Business data lineage
Overview
Profile
Data Quality
Data Preview
Lineage
Relationships
Data profiling
Date of birth
Email
Contact
SIN
Gender
First name
Insights
Anomalies detected in attribute birth_date
NULL , N/A , and 1900-01-01 are outliers.
Apply automatic cleansing before using this data.
3 outlier values
detected in given attribute
Anomaly detection
Data Quality
Last evaluated 1 hours ago
Overall Validity
67%
3
Records
1 230 954
453
Checks
4
Attributes
345
40
Data quality
Unexpected length and record count
Detected based on previous 14 versions of the profile
Open detail
Name
Terms
src_primary_key
first_name
First name
birth_date
Date of birth
gender_id
Gender
Relations
Parent of
Governed by
GDPR policy
Transformed items
Party_full
Relationship discovery