Data Quality and Observability

Building a Trusted Foundation for Analytics and AI Excellence: Optimizing Data Quality

Seán O' Shea ,

Senior Manager , Product Management

Disclaimer: The information being provided herein is for informational purposes only. The development, release and timing of any Informatica product, service or functionality described herein remain at the sole discretion of Informatica and should not be relied upon in making a purchasing decision. Statements made herein are based on information currently available, which is subject to change. Such statements should not be relied upon as a representation, warranty or commitment to deliver specific products, services or functionality in the future.
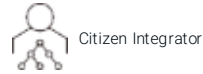
AI-Powered Cloud Data Management

DATA CONSUMERS

ETL Developer   Data Engineer   Citizen Integrator   Data Scientist   Data Analyst   Business Users

Intelligent Data Management Cloud™

DISCOVER & UNDERSTAND	ACCESS & INTEGRATE	CONNECT & AUTOMATE	CLEANSE & TRUST	MASTER & RELATE	GOVERN & PROTECT	SHARE & DEMOCRATIZE
DATA CATALOG	DATA INTEGRATION & ENGINEERING	API & APP INTEGRATION	DATA QUALITY & OBSERVABILITY	MDM & 360 APPLICATIONS	GOVERNANCE, ACCESS & PRIVACY	DATA MARKETPLACE

CLAIRE®
AI-Powered Metadata Intelligence & Automation

Connectivity
Metadata System of Record

DATA SOURCES

SaaS Apps Sources   +   Mainframe   Applications   Databases   On-premises Sources   +   IoT   Machine Data   Logs   Real-time / Streaming Sources

Informatica

The Industry's Only **Intelligent Data Management Cloud**

Open and Flexible

Designed to work within your reference architecture

Low Code, No Code

Increased productivity for data engineers and citizen integrators

Best of Breed

Industry leading products and services all in one unified platform

Multi-cloud and Hybrid

Runs on and interoperates with what you have today and tomorrow

Leverages Best of Breed Open Source

Delivers open-source innovations without the complexity

API and Microservices Based

Modern architecture for optimal performance and resiliency

Elastic and Serverless

Cost optimized performance and scalability

Cloud Native

Built from the ground up for enterprise cloud workloads

Secure

Highest level of cloud security certifications and attestations

Single Uniform Consumption Pricing

Predictive and flexible pricing that adjusts to just what you need today

Informatica

Compliance
Are we complying with all regulatory reporting requirements?

Sales
How do I increase win rates, upsell and cross-sell?

Marketing
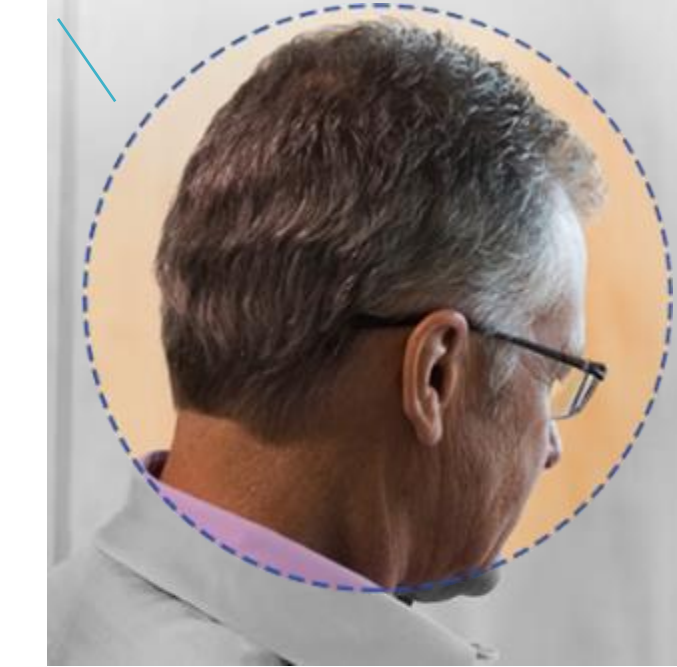Can I improve customer engagement and conversion?

Data Scientist
How can I discover, understand and trust data required for my analysis?

IT
How do we minimize the risk of migrating to the cloud?

Business User
Is our data ready to support our digital transformation initiatives?

Put Simply
Garbage in, Garbage out

Garbage Data

AI
Analytics
Data Pipelines
Migrations
Modernization
…

Garbage Results

Informatica

Data Quality

High-quality data is the essential ingredient for high-quality analytics.
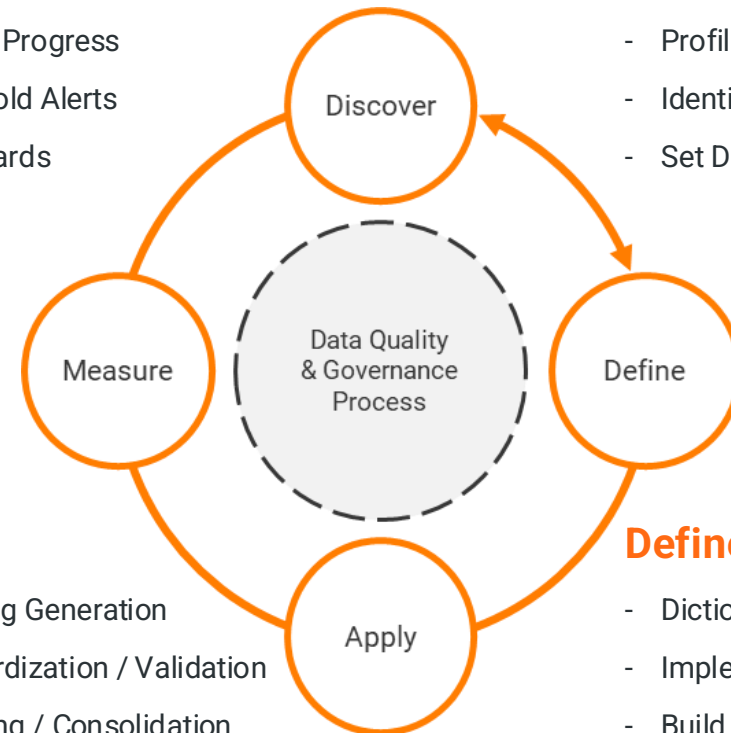
What is Data Quality ?

- There are many definitions of data quality!

  - trustworthy & accurate

- Two Key characteristics

  - **High Quality Data**

  - **Process by which to achieve high quality data**

  (improve & monitor it)

Measure
- Review Progress
- Threshold Alerts
- Scorecards

Discover
- Profiling
- Identify Data Issues
- Set Data Quality Goals

Apply
- Mapping Generation
- Standardization / Validation
- Matching / Consolidation

Define
- Dictionaries
- Implement Rules
- Build Cleanse, Parse, Verification, etc. processes

Discover

Measure

Data Quality & Governance Process
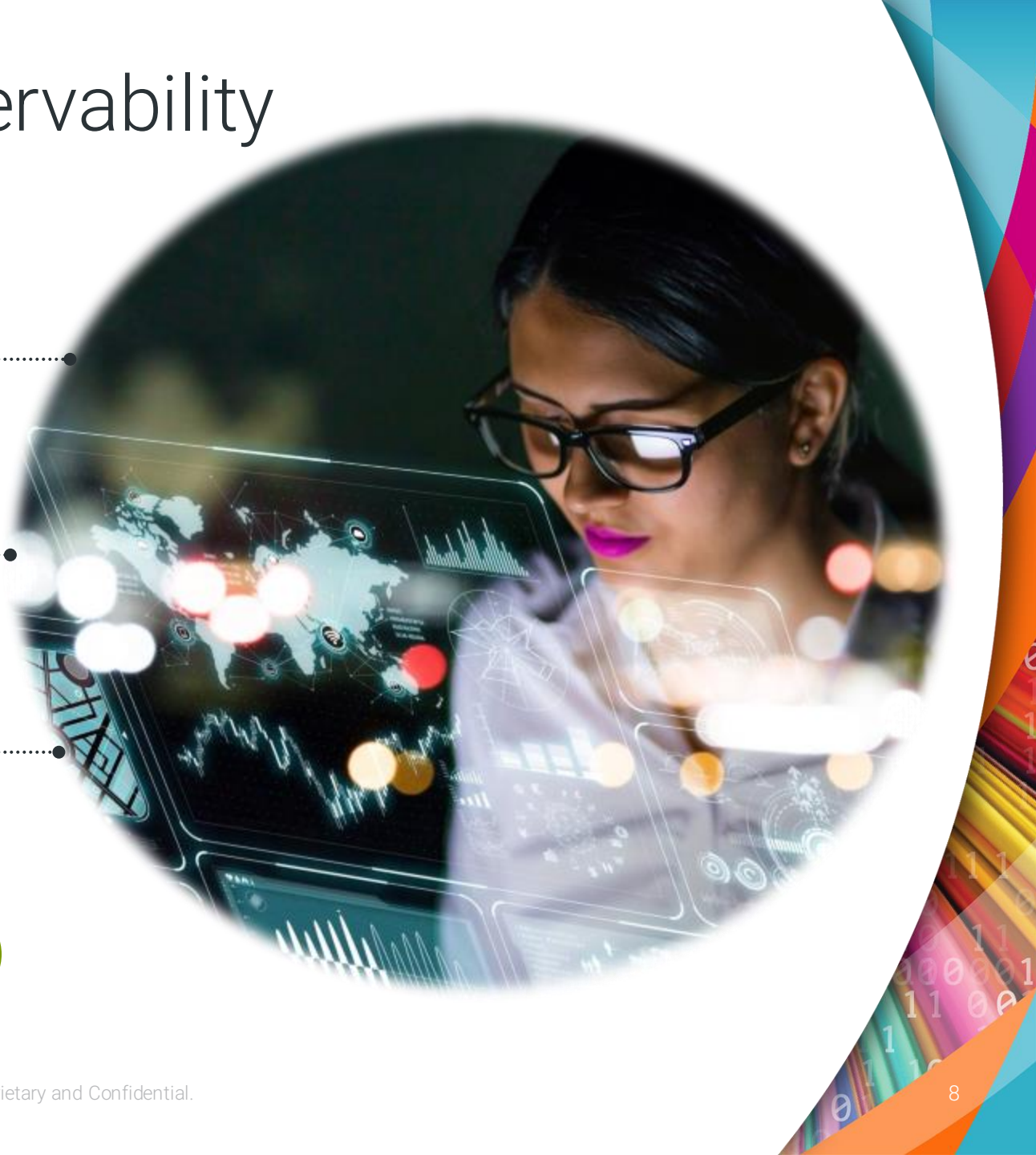
Define

Apply

Cloud Data Quality and Observability

Operationalize trust

Persona-aligned views on health

Automation for speed and scale

Manage and scale high-quality data pipelines for analytics & AI

Informatica

Data - Provide Visibility Into Data Health

Proactive detection of data quality issues and anomalies for faster remediation

- Proactively Monitor Data Quality

  - Ensure that data is complete, accurate, and consistent and pinpoint the source of issues and the potential impact

- Detect Anomalies and Outliers

  - Highlight data points that are unusual or unexpected and instances where data violates business rules and create alerts

  - Observe changes in various metrics of the data and alert data owners about anomalies in the changes

  - Learn from user input to automatically adjust anomaly detection algorithms over time

Informatica

DEMO

Data Quality

Generic Primary Org
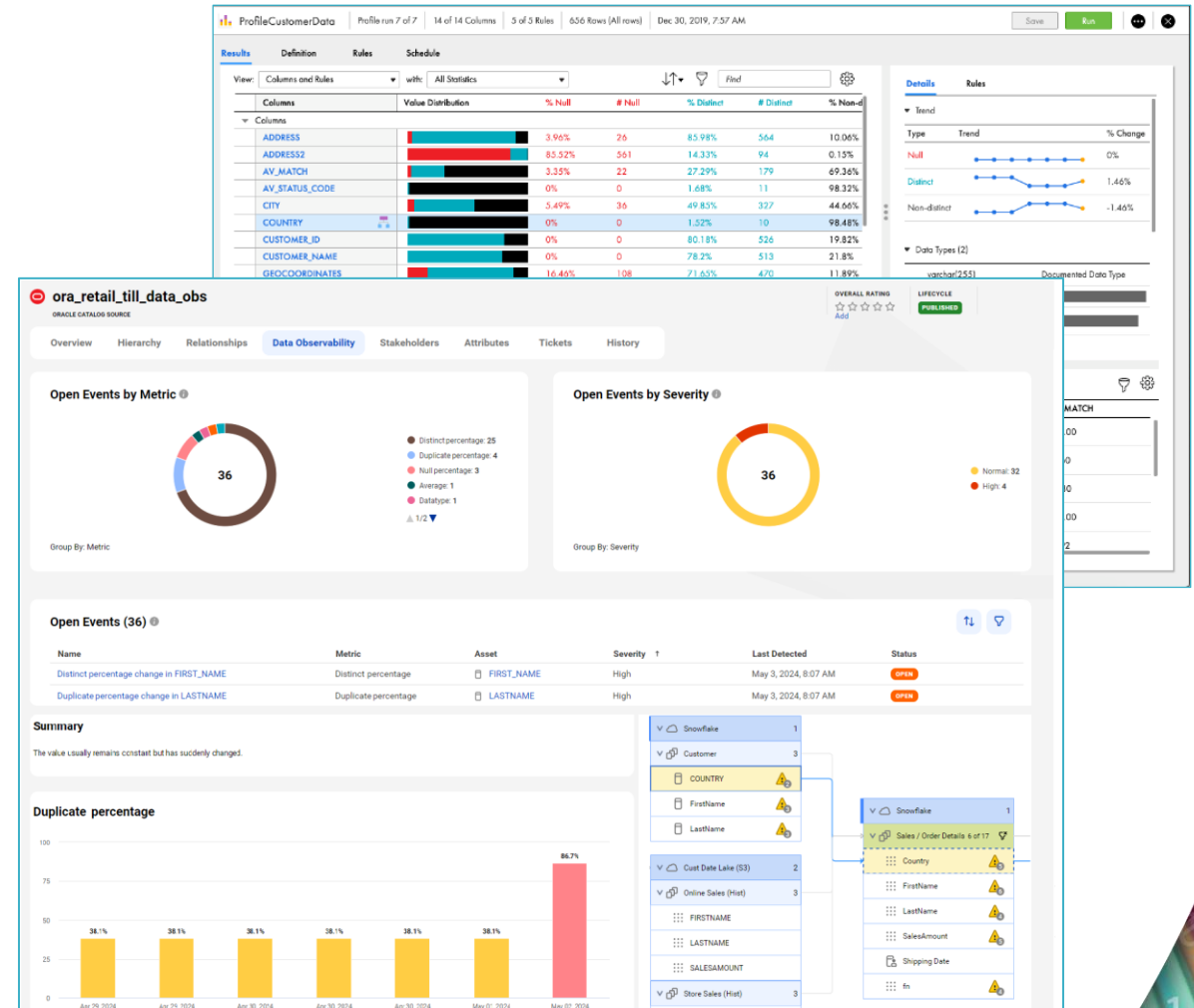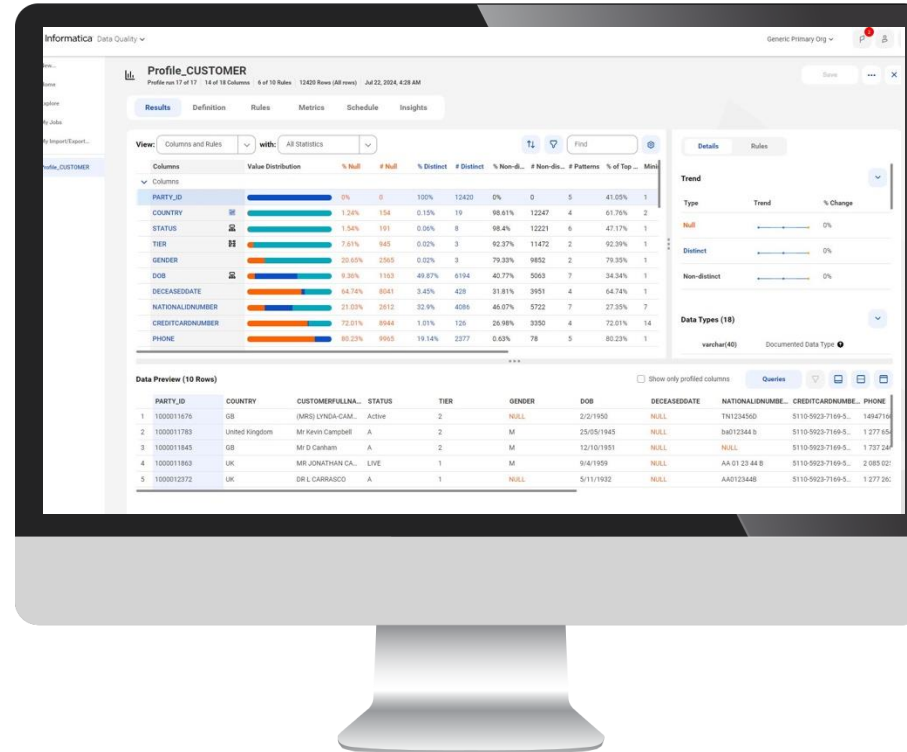
Profile_CUSTOMER

Profile run 17 of 17 | 14 of 18 Columns | 6 of 6 Rules | 12420 Rows (All rows) | Jul 22, 2024, 4:28 AM

Save | Run

Results | Definition | Rules | Metrics | Schedule | Insights

View: Columns and Rules | with: All Statistics | Find

Columns	Value Distribution	% Null	# Null	% Distinct	# Distinct	% Non-di...	# Non-dis...	# Patterns	% of Top ...	Min
Columns										
PARTY_ID		0%	0	100%	12420	0%	0	5	41.05%	1
COUNTRY		1.24%	154	0.15%	19	98.61%	12247	4	61.76%	2
STATUS		1.54%	191	0.06%	8	98.4%	12221	6	47.17%	1
TIER		7.61%	945	0.02%	3	92.37%	11472	2	92.39%	1
GENDER		20.65%	2565	0.02%	3	79.33%	9852	2	79.35%	1
DOB		9.36%	1163	49.87%	6194	40.77%	5063	7	34.34%	1
DECEASEDDATE		64.74%	8041	3.45%	428	31.81%	3951	4	64.74%	1
NATIONALIDNUMBER		21.03%	2612	32.9%	4086	46.07%	5722	7	27.35%	7
CREDITCARDNUMBER		72.01%	8944	1.01%	126	26.98%	3350	4	72.01%	14
PHONE		80.23%	9965	19.14%	2377	0.63%	78	5	80.23%	1

Details | Rules

Trend

Type	Trend	% Change
Null		-34.97%
Distinct		0%
Non-distinct		34.97%

Data Types (1)

varchar(40) | Documented Data Type

## Data Preview (10 Rows)

Show only profiled columns | Queries

	PARTY_ID	COUNTRY	CUSTOMERFULLNA...	STATUS	TIER	GENDER	DOB	DECEASEDDATE	NATIONALIDNUMBE...	CREDITCARDNUMBE...	PHONE
1	1000011676	GB	(MRS) LYNDA-CAM...	Active	2	NULL	2/2/1950	NULL	TN123456D	5110-5923-7169-5...	1494716
2	1000011783	United Kingdom	Mr Kevin Campbell	A	2	M	25/05/1945	NULL	ba012344 b	5110-5923-7169-5...	1 277 654
3	1000011845	GB	Mr D Canham	A	2	M	12/10/1951	NULL	NULL	5110-5923-7169-5...	1 737 240
4	1000011863	UK	MR JONATHAN CA...	LIVE	1	M	9/4/1959	NULL	AA 01 23 44 B	5110-5923-7169-5...	2 085 025
5	1000012372	UK	DR L CARRASCO	A	1	NULL	5/11/1932	NULL	AA012344B	5110-5923-7169-5...	1 277 262
6	1000012381	UK	MR A CARRIER	A	2	M	2/5/1961	NULL	QD012344C	5110-5923-7169-5...	NULL

Collapse

Start typing search terms for suggestions.

Generic Primary Org ⌄

+ New...

🏠 Home

◎ Browse

🔍 Saved Searches

📊 Profile_CUSTOMER

Cloud DQ Standard Demo / @Profile / **Profile_CUSTOMER**   ✕

Average Latest Scores by Dimensions

Validity (1)

99.87%

↑ 0%

Completeness (2)

94.69%

↑ 2.21%
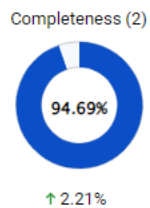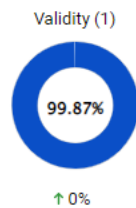
Number of Rule Occurrences by Dimensions

● Good
● Acceptable
● Not Acceptable

Validity

Completeness

0   1   2

Values

Rule Occurrences (3)   ⇅ ▽ [Find] ⚙

Name	Dimension	Latest Score	Date	Total Rows	Failed Rows	Primary Data Element	Description
VALIDITY_DOB	Validity	🟡 99.87%	Jul 22, 2024, 4:30 AM	12420	15	DOB	Consistency of DOB bas...
DOB_Completeness	Completeness	🔴 90.63%	Jul 22, 2024, 4:30 AM	12420	1163	DOB	Determines whether an ... ⋯
DOB_Completeness ...ness	Completeness	🟡 98.76%	Jul 22, 2024, 4:30 AM	12420	154	COUNTRY	Completeness of COUN...

Data Quality for Everyone, Everywhere

Use Cases

Analytics, AI and ML	Data Engineering	Master Data Management	Data Governance	Data Observability
Profiling	Profiling	Profiling	Profiling	Anomaly Detection
Matching	Matching	Parsing	Auto Rule Generation	Data Lineage
Standardization and Cleansing	Parsing	Standardization and Cleansing	Standardization and Cleansing	Performance Optimization
Verification and Enrichment	Standardization and Cleansing	Verification and Enrichment	Data Quality Metrics	Monitor and Alerting

Core Capabilities Needed

Multidomain – Scalable – Hybrid

Data Analyst   Data Scientist   Data Engineer   Data Architects   Line of Business   Chief Data Officer   Data Ops

Informatica
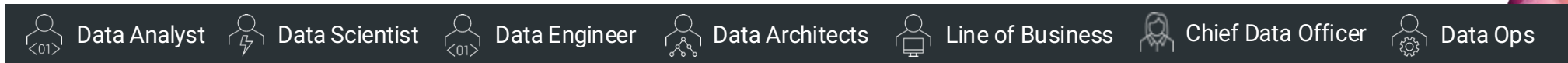
Thank You

Seán O' Shea / September 18th, 2024

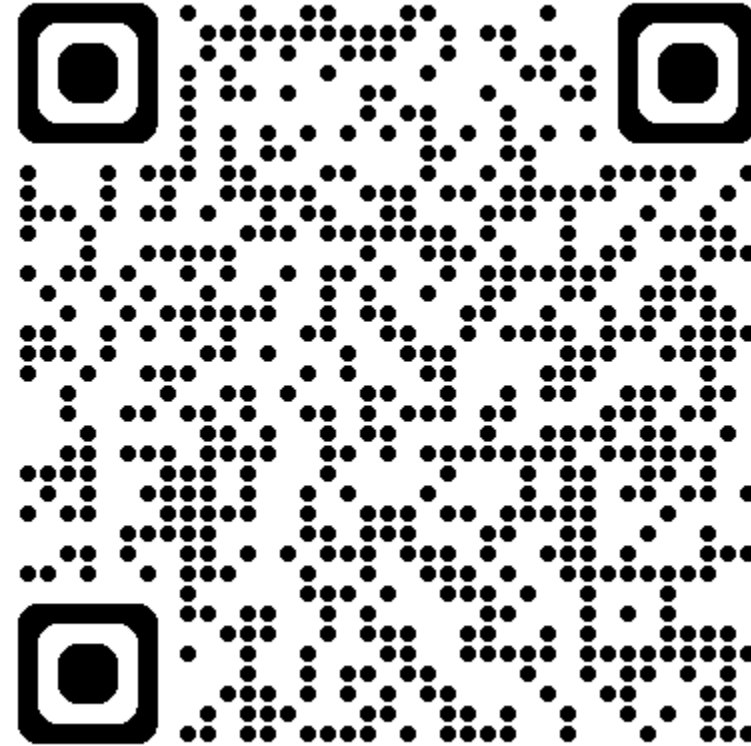**Where data & AI come to LIFE**

Free 30-Day Trial: Cloud Data Quality

Experience a no-code, visual environment for creating and managing data quality capabilities—all built on a leading next-generation iPaaS.

In this trial, you can:

- ✅ Create business logic to validate assumptions and statements about your data.

- ✅ Verify the accuracy and formatting of data values on a data source or another object in a mapping.

- ✅ Publish and consume reusable data quality artifacts across all data integration initiatives, including cloud data warehousing and cloud data migrations.

https://www.informatica.com/trials/cloud-data-quality.html

Where data & AI come to **LIFE**