A use case in solving data quality management at scale

Tom Loxton

20 August 2024

An urgent need to implement robust data quality management

Mounting regulatory pressure...

...in a complex Financial Institution

An increasing level of regulatory pressure on the organisation - partly related to what is called "Data Issues"

A European financial institution with $14B market cap, operating in over 15 countries, 40 subsidiaries, 3,000 data sources, 24 000 employees & 6 million customers

Data is key to achieve strategic goals, gain competitive advantage in the market and ensure business continuity in adverse situations. But in order to use it to its full potential we need to consider the trustworthiness of the information we have at hand and to be able to execute corrective measures when is not up to standards.

PREVENT, DETECT, CORRECT
Establishing Data Quality Management requires changes to Processes, Technology & Organizational structures

Corrective and detective actions provide feedback on which preventive measures should be strengthened

Corrective actions provide feedback on which detective measures should be put in place

PREVENT

*Learn and adapt*

DETECT

Learn and adapt

CORRECT

With preventive measures aim to avoid data issues to appear in the first place – ensuring we create fit-for-purpose from the start. This requires Data Management capabilities, but more often other business capabilities

Data elements are monitored on Data Quality criteria – ensuring we have an early warning system signalling before Data Quality starts to deteriorate. Empowering the organisation to swiftly act

If data issues start to appear lean processes are in place to facilitate corrective action – ensuring we cleanse, remediate and close the tap timely in a cost-effective manner

- Prevention measures lie in the Data Management domain but **heavily relies on strengthening other business capabilities** as well. Therefore, collaboration in these areas needs to be strengthened to take the necessary steps

- Examples outside the Data Management where improvement will lead to preventing data issues:
  - Source System Change Management
  - Architecture (control)
  - IT development
 - Business Process Management

Technology
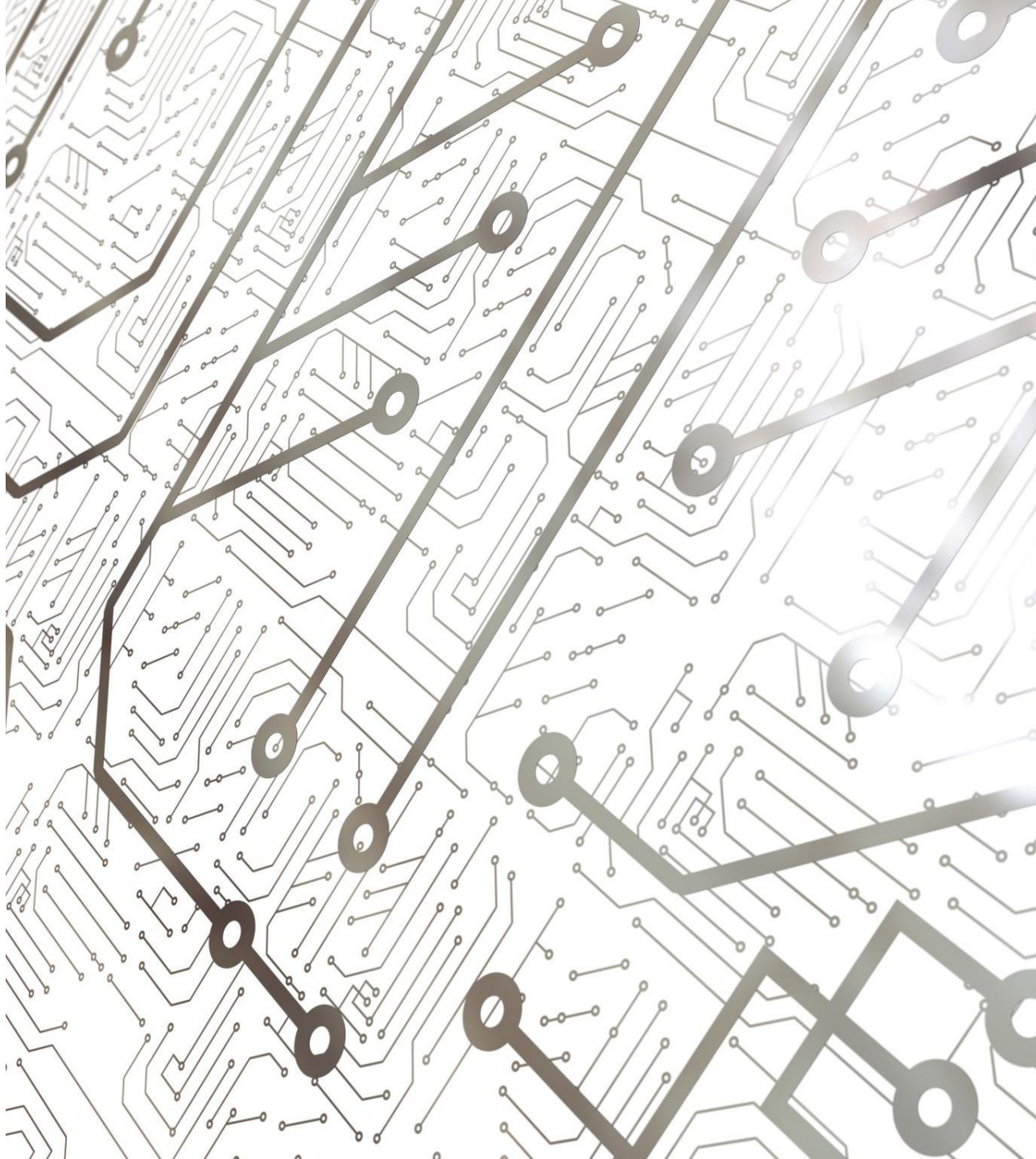
Data Quality Monitoring

is a capability that enables Data Owners to create continuous transparency on the trustworthiness of data and to proactively act when it does not meet the expectations agreed between Data Users and Data Owners.

PREVENT → *Learn and adapt* → DETECT → *Learn and adapt* → CORRECT

How does it work?	Product developed	
Data Users defines **DQ Requirements** and reaches agreements on expectations with Data Owners	**DQ Requirements Manager**	
Data Owner defines **DQ Checks** on their Golden Data Elements to monitor the fulfilments of the agreed expectations	**DQ Check Builder**	
DQ Checks results are used to calculate **DQ Scores** and measure trustworthiness of data	**DQ Scores**	
Detected **DQ Signals** (issues) are routed to cleansing parties/processes for correction	**DQ Signal Manager**	
If the amount of signals on certain data element surpasses a threshold, an **automated DQ Issue** is raised root cause investigation.	**DQ Issues auto creation**	

To create fit-for-purpose, rapidly accessible data, the *Data Marketplace* was built.
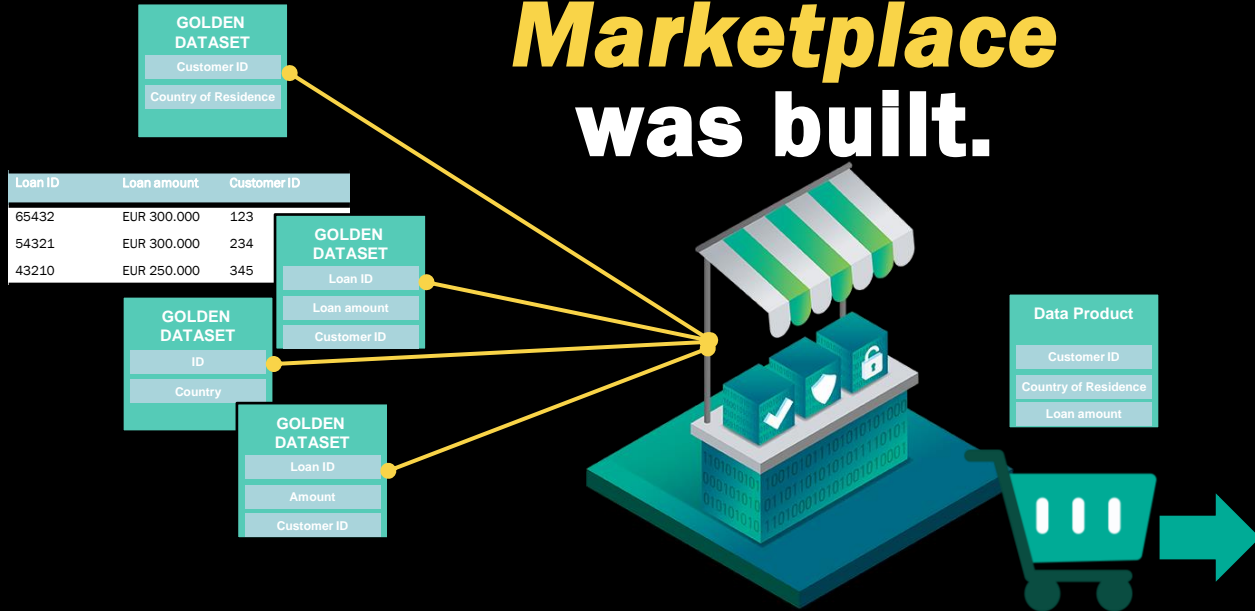
Data Quality Scores are available in the Data Marketplace for all Golden Datasets and Data Products

GOLDEN DATASET
- Customer ID
- Country of Residence

| Loan ID	Loan amount	Customer ID
65432	EUR 300.000	123
54321	EUR 300.000	234
43210	EUR 250.000	345

GOLDEN DATASET
- Loan ID
- Loan amount
- Customer ID

GOLDEN DATASET
- ID
- Country

GOLDEN DATASET
- Loan ID
- Amount
- Customer ID

Data Product
- Customer ID
- Country of Residence
- Loan amount

← back   Preview in sidebar

Money Mule Prediction Integrated Dataset

Customer data of natural persons and legal entities ABN AMRO Retail Banking. Customer data of natural persons and legal entities ABN AMRO Retail Banking. Customer data of natural persons and legal entities ABN AMRO Retail Banking.

Daan Oona — Data User   View →
Steward de Vries with a long name — Data Steward
14 Application Owners Consuming Applications

CIA 1 2 3
Innovation & Technology
IDS_R2
14 Applications
- 4 Golden Sources
- 4 Consuming Applications
- 6 Other   Other Applications in the chain
3 Integrated Data Elements
28 Golden Data Elements
2 Critical Data Elements
No personal data
Normal personal data
Sensitive personal data
Particularly sensitive personal data

● Data Quality Score
70% Current
80% (Last)
90% (Threshold)

Data Quality Issues
20 15 10 5
2
Feb Mar Apr May

Data Gaps
20 15 10 5
5
Feb Mar Apr May

Golden Datasets   15   5
Data Sharing Agreements   14   This month   7
Use Cases   9   2

Integrated Data Elements (3)   Find

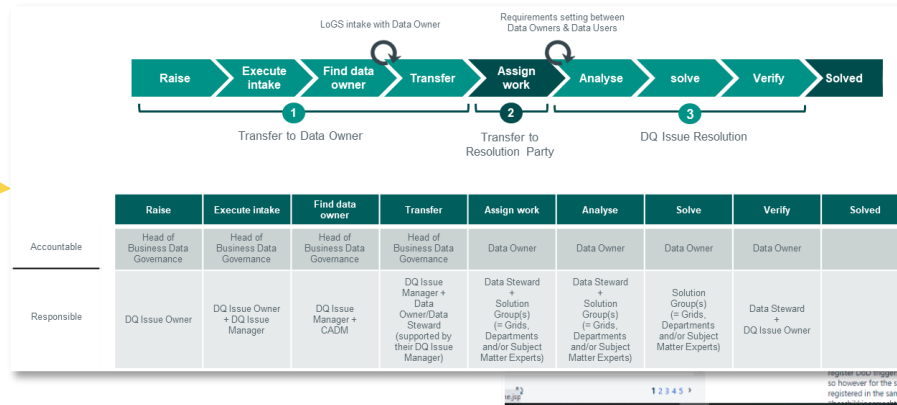| Name ▲	Term & Definition ▼	Label ▼	DQ Score ▼	DQ Issues ▼	Golden Data Elements ▼	Relationships ▼
1 **Official Name** OFF_NAME	Official name Official name as stated on the identity card of the natural person. The official name consists... *more*	CDE	70%	24 ▲ 1	16	
2 **Postcode** POST_AREA RANGE_CODE	Postal Code A short sequence of characters as part of a home or postal address.	Potentially Sensitive	92%	24 ▼ -1	n/a	
3 **Official Name** OFF_NAME INTERFACE	Official name Official name as stated on the identity card of the natural person. The official name consists... *more*	CDE	85%	2	ID123456 3 Data Gap(s) 12 6 mapped 3 validated mapping	

Processes
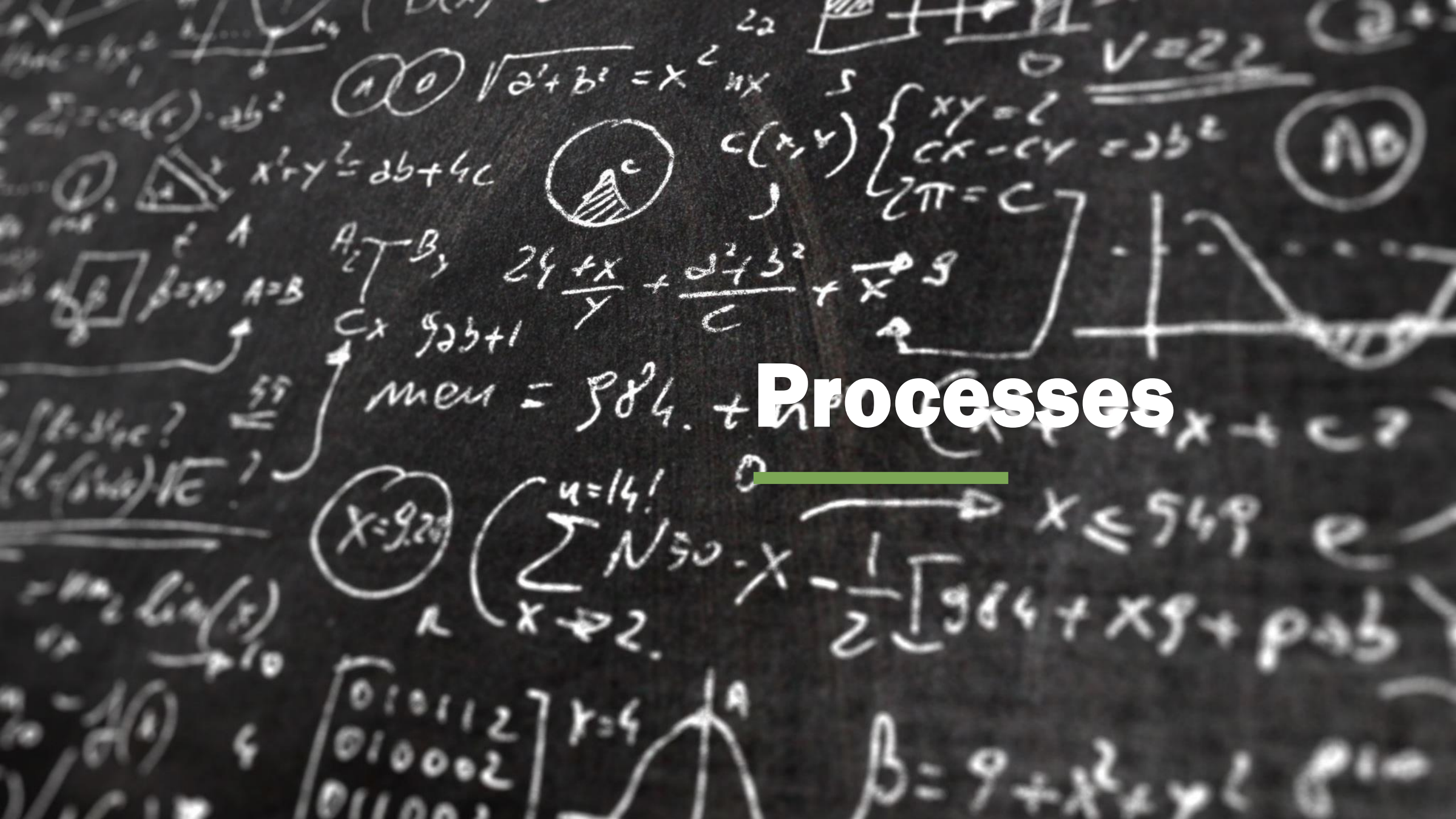
Data Issue Management: To make fit-for-purpose data, we embedded a Data Issue Management process in the bank – ensuring data quality issues are managed and structurally resolved

PREVENT — *Learn and adapt* — DETECT — *Learn and adapt* — CORRECT

ENCOURAGE ANYONE TO RAISE DATA QUALITY ISSUES

ABN AMRO / DQ Issue Service Desk
Data Quality Issue E-form

Raise a new DQ issue / Meld een nieuw DQ issue

Is your data inaccurate (or incorrect), incomplete, inconsistent and/or not timely?
Then raise your DQ issue via this form (preferably in English).

Is jouw data niet accuraat (incorrect), niet compleet, inconsistent of niet tijdig?
Meld dan jouw DQ issue via dit formulier (bij voorkeur in het engels).

A ONE BANK-WIDE PROCESS FOR TRANSPARANCY – GUIDING THE ISSUES TOWARDS CLOSURE

LoGS intake with Data Owner

Requirements setting between Data Owners & Data Users

Raise → Execute intake → Find data owner → Transfer → Assign work → Analyse → solve → Verify → Solved

Transfer to Data Owner | Transfer to Resolution Party | DQ Issue Resolution

	Raise	Execute intake	Find data owner	Transfer	Assign work	Analyse	Solve	Verify	Solved
Accountable	Head of Business Data Governance	Head of Business Data Governance	Head of Business Data Governance	Head of Business Data Governance	Data Owner	Data Owner	Data Owner	Data Owner	
Responsible	DQ Issue Owner	DQ Issue Owner + DQ Issue Manager	DQ Issue Manager + CADM	DQ Issue Manager + Data Owner/Data Steward (supported by their DQ Issue Manager)	Data Steward + Solution Group(s) (= Grids, Departments and/or Subject Matter Experts)	Data Steward + Solution Group(s) (= Grids, Departments and/or Subject Matter Experts)	Solution Group(s) (= Grids, Departments and/or Subject Matter Experts)	Data Steward + DQ Issue Owner	

View all issues and filters

1 of 5859

⌄ People
Assignee: Robert van Rossum
Reporter: Rob Hendriks
Request participants: None
Organizations: None
Votes: Vote for this issue
Watchers: Start watching this issue

⌄ Dates
Created: 5 hours ago

LINKED TO REGULATORY THEMES

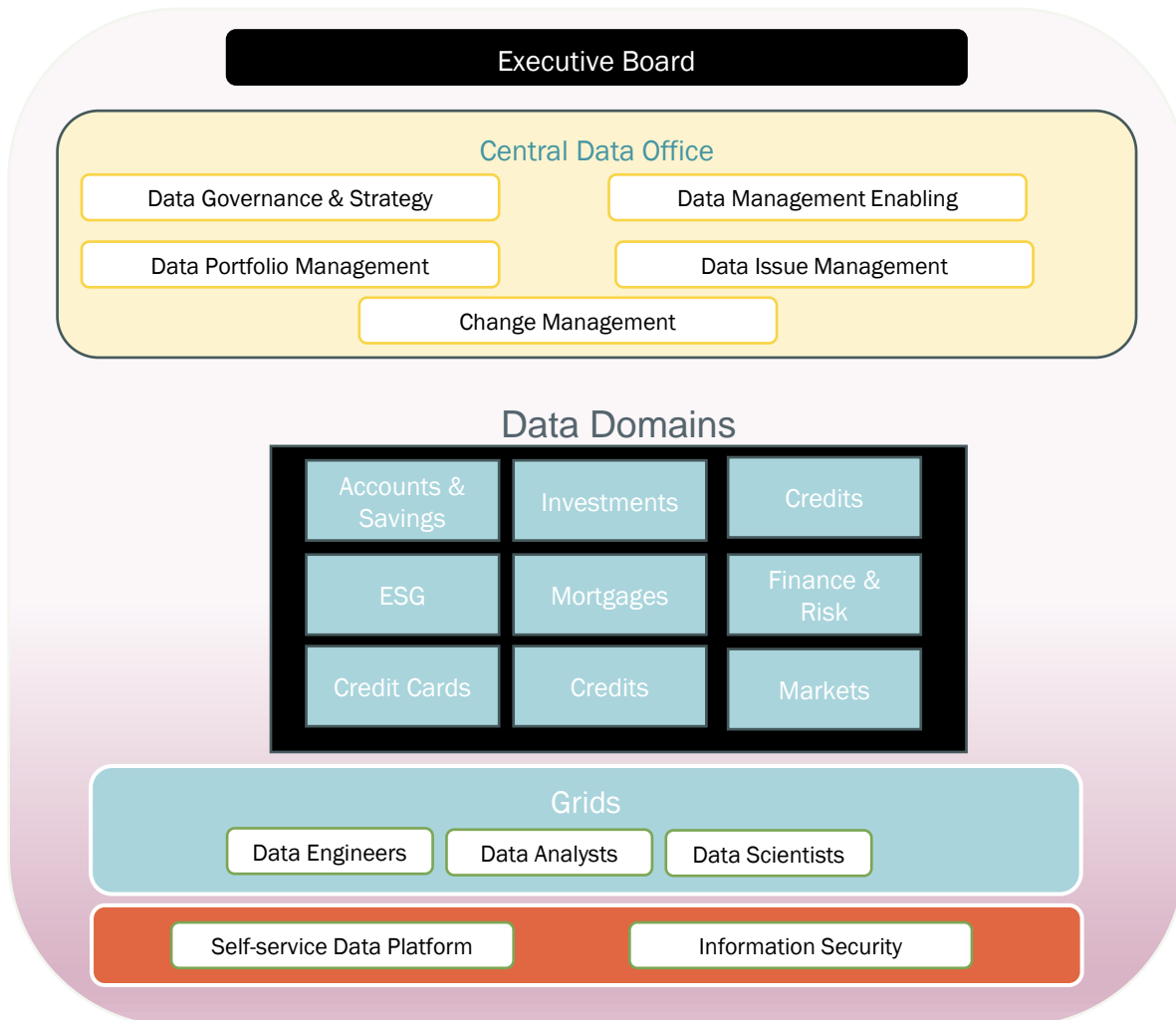| Theme	Planning status of open DQ issues 2021				Closed DQ issues 2021			
	Planned for resolution	In planning Q4²	Remaining	Total	Solved	Cancelled – Data Gaps	Cancelled – Other	Total
Future Model Landscape (Theme-24)	21	40	152	213	40	6	550	596
FRAAI (Theme-23)	155	75	75	305	92	19	65	176
Detecting Financial Crime (Theme-22)	61	8	11	80	10	0	2	12
Sub total unique DQ Issues for Cluster 1¹	207 (37%)	116 (21%)	237 (42%)	560	140	25	617	782
Future Credit Domain (Theme-51)	42	18	33	93	4	3	13	20
Substantable Finance Regulation (Theme-16)	3			3				
IFRS9 Impairments (Theme-13)	8	1	23	32	4		37	41
Granular Reporting (Theme-12)	55	32	55	142	25	13	19	57
Duty of Care Retail – Customised (Theme-9)	2			2				
Deposit Guarantee Scheme (Theme-7)	30	6	1	37	3		3	6
CRR2/CRD5 Reporting & Disclosures (Theme-4)		1		1	1		1	1
Basel III/IV (Theme-1)	24	1	6	34	9	2	7	18
Sub total unique DQ Issues for Cluster 2¹	134 (44%)	58 (19%)	114 (37%)	306	46	14	72	132
Other	45	11	320	376	96	3	250	349
Total unique issues¹	264 (26%)	130 (13%)	610 (61%)	1004¹	242	28	905	1175¹

Data Issues are raised by anyone encountering an issue. Correcting issues is coordinated via a process and prioritised in line with timelines of Regulatory themes
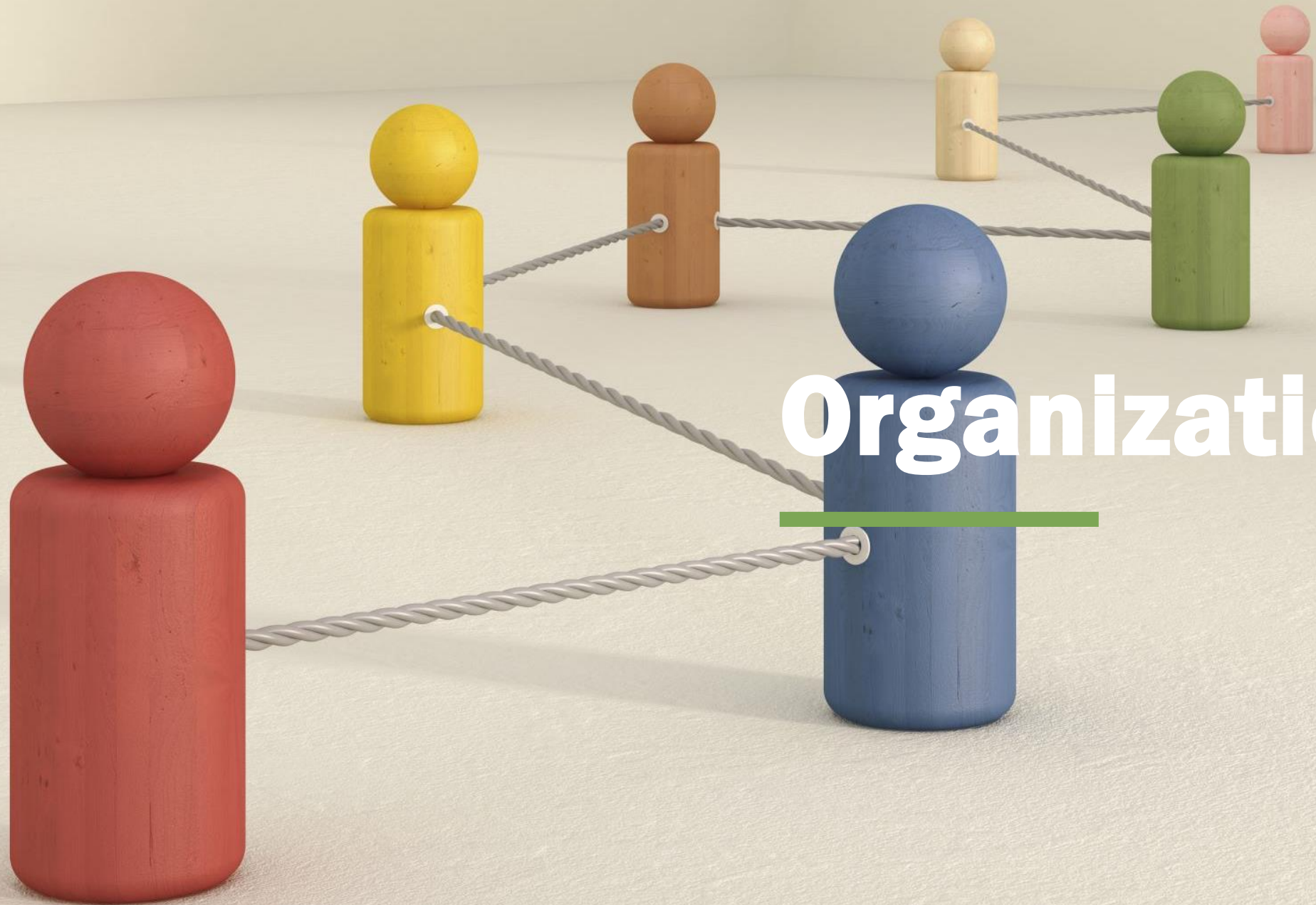
Organization

To embed Data Management across the bank a decisive data organisation is set up using a *Federated Data Governance Model*, adapted from Data Mesh principles

In a data mesh approach, the organizational structure shifts from a centralized data team to a decentralized model where data ownership and responsibility are distributed across domain-specific teams. Each team, aligned with a particular business domain (e.g., marketing, finance, sales), acts as a "data product" owner, responsible for the creation, maintenance, and governance of their domain's data products.

Executive Board

Central Data Office

Data Governance & Strategy

Data Management Enabling

Data Portfolio Management

Data Issue Management

Change Management

Data Domains

Accounts & Savings	Investments	Credits
ESG	Mortgages	Finance & Risk
Credit Cards	Credits	Markets

Grids

Data Engineers

Data Analysts

Data Scientists

Self-service Data Platform

Information Security

Data quality management is decentralized, with each domain team taking ownership of the quality of their own data products.

Domain Ownership

- **Responsibility:** Each domain team is accountable for the quality of their data products. This includes ensuring data accuracy, completeness, consistency, and timeliness. Since these teams are closest to the source and use of the data, they are best positioned to define and enforce quality standards that meet the specific needs of their domain.

Embedded Data Quality Practices

- **Quality by Design:** Data quality is embedded into the data product development process from the outset. Teams define clear quality metrics and use automated tools to monitor and maintain these standards continuously.

Centralized Support & Standards

- **Platform Team:** A centralized platform engineering team provides the necessary infrastructure, tools, and guidelines to support data quality management across all domains. This includes providing data quality monitoring tools, standardized metrics, and shared best practices.

- **Global Standards:** While domain teams are autonomous, they adhere to overarching organizational standards for data quality, interoperability, and compliance. This ensures that data products can be effectively integrated and used across the organization