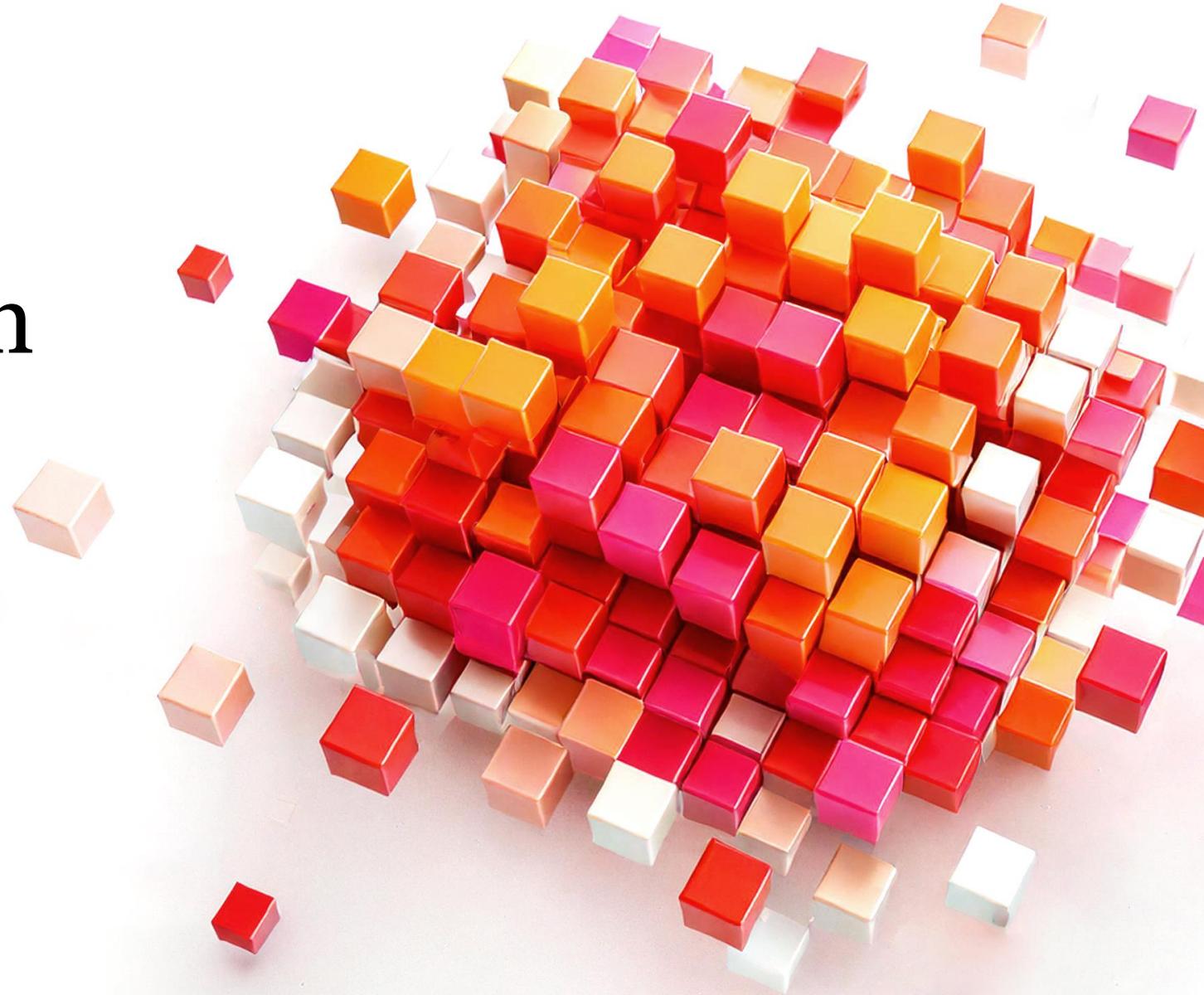Data Day 2025

strategy&
Part of the PwC network

Leap into Next Gen Data Management

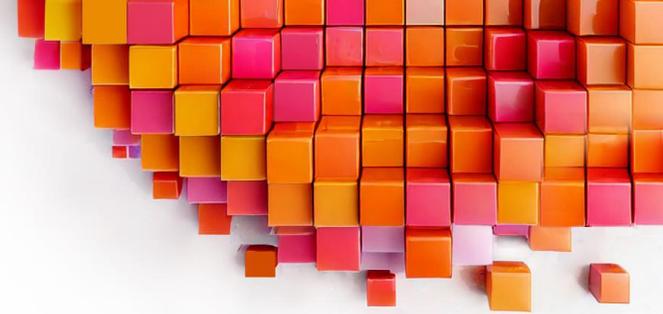How to tackle challenges and harness cutting-edge technology for data management?
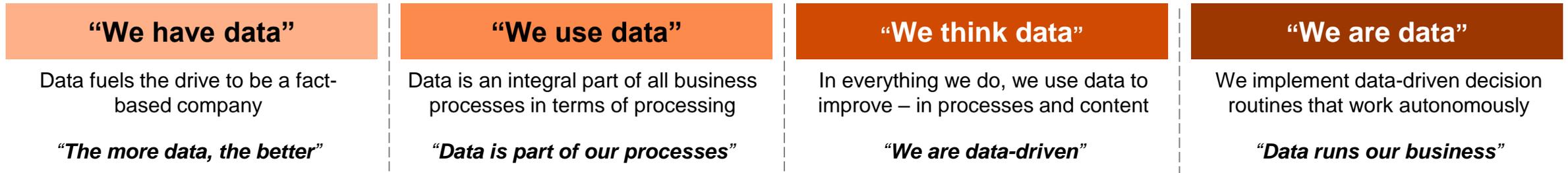
pwc

01 How do we overcome typical challenges with operationalizing Data Management?

We expect data capabilities to become a differentiator for all industries

PwC Data Maturity Model

"We have data"	**"We use data"**	**"We think data"**	**"We are data"**
Data fuels the drive to be a fact-based company	Data is an integral part of all business processes in terms of processing	In everything we do, we use data to improve – in processes and content	We implement data-driven decision routines that work autonomously
"The more data, the better"	*"Data is part of our processes"*	*"We are data-driven"*	*"Data runs our business"*

Common topics & challenges on the way to becoming a data-driven organization

Management commitment	Degree of centralization	Data sharing	Data access
Data products	End-to-end responsibilities	Data Culture	People & organization

Data Domains are the key enabler for increasing value-based data usage

Use-Cases
Analytics & AI use-cases along the value chain

R&D	Procurement	Supply Chain	Manufacturing	Sales	Finance
• Product Design & Simulation • AI based testing	• Procurement assistent • Supplier rating	• Advanced Forecasting • Inventory Opt.	• Predictive Maintenance • Predictive QUA	• Customer 360° • Next best Action	• Cashflow Forecasting • Auto Commenting

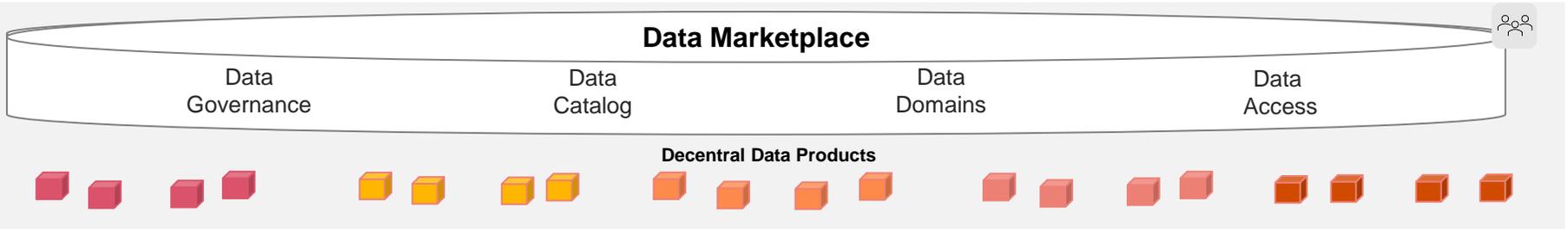**Cross use-cases**   Sustainability Control Tower   Digital Twin   Supply Chain Due Dilligence Act   Product Passport

Data Marketplace Components
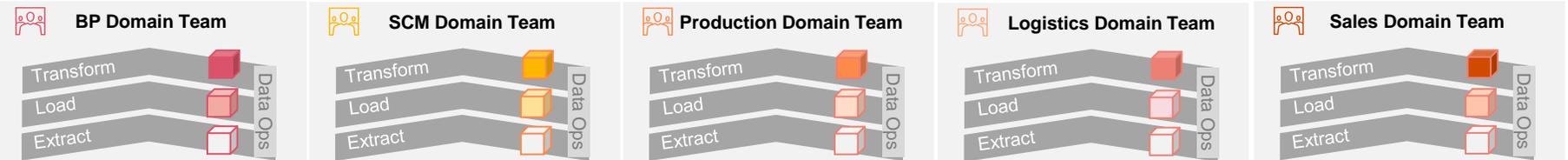Four central services allows for effectively handling of the decentrally available data products

Data Marketplace

Data Governance   Data Catalog   Data Domains   Data Access

Consumable Data Products
Each domain creates and shares valuable data products

Decentral Data Products

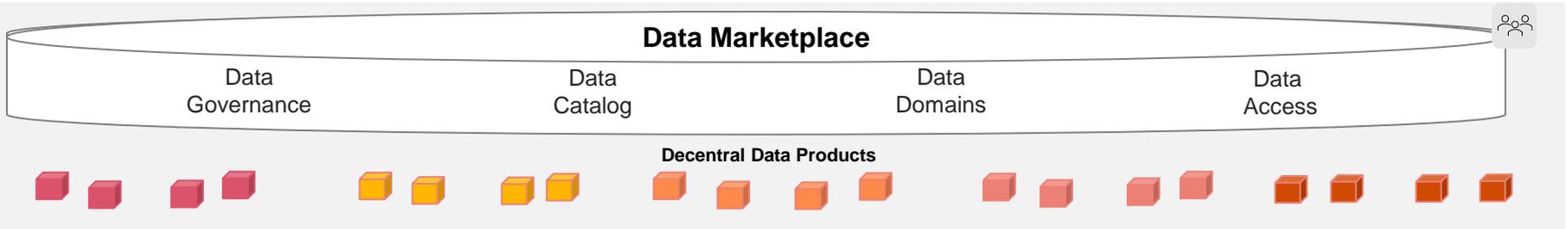**Interdisciplinary Domain Teams**
Data is owned by teams in federated domains that business structure

BP Domain Team	SCM Domain Team	Production Domain Team	Logistics Domain Team	Sales Domain Team
Transform Load Extract Data Ops	Transform Load Extract Data Ops	Transform Load Extract Data Ops	Transform Load Extract Data Ops	Transform Load Extract Data Ops

Digital Footprint
Operational Data and IoT Data that reflect real-world supply chain processes

BP Domain Data	Warehouse Domain Data	Production Domain Data	Distribution Domain Data	Customer Domain Data
SRM Data / IoT Data	WMS Data / IoT Data	MES Data / IoT Data	TMS Data / IoT Data	POS Data / IoT Data

Physical World
Movement, storage and manufacturing of physical goods

Supplier → Warehouse → Production → Distribution → Customer

Data Domains are the key enabler for increasing value-based data usage

Use-Cases
Analytics & AI use-cases along the value chain

R&D
- Product Design & Simulation
- AI based testing

Procurement
- Procurement assistent
- Supplier rating

Supply Chain
- Advanced Forecasting
- Inventory Opt.

Manufacturing
- Predictive Maintenance
- Predictive QUA

Sales
- Customer 360°
- Next best Action

Finance
- Cashflow Forecasting
- Auto Commenting

Cross use-cases   Sustainability Control Tower   Digital Twin   Supply Chain Due Dilligence Act   Product Passport

Data Marketplace Components
Four central services allows for effectively handling of the decentrally available data products
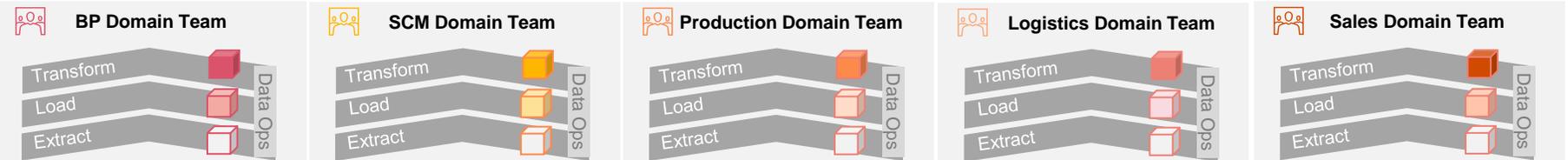
Data Marketplace

| Data Governance | Data Catalog | Data Domains | Data Access |

Consumable Data Products
Each domain creates and shares valuable data products

Decentral Data Products

Interdisciplinary Domain Teams
Data is owned by teams in federated domains that business structure

BP Domain Team
- Transform
- Load
- Extract
- Data Ops

SCM Domain Team
- Transform
- Load
- Extract
- Data Ops

Production Domain Team
- Transform
- Load
- Extract
- Data Ops

Logistics Domain Team
- Transform
- Load
- Extract
- Data Ops

Sales Domain Team
- Transform
- Load
- Extract
- Data Ops

Digital Footprint
Operational Data and IoT Data that reflect real-world supply chain processes

BP Domain Data
- SRM Data
- IoT Data

Warehouse Domain Data
- WMS Data
- IoT Data

Production Domain Data
- MES Data
- IoT Data

Distribution Domain Data
- TMS Data
- IoT Data

Customer Domain Data
- POS Data
- IoT Data

Physical World
Movement, storage and manufacturing of physical goods

Supplier → Warehouse → Production → Distribution → Customer

Top management commitment is key to shifting the mindset around data access

Need-to-know

1	**Starting point** – All data is inaccessible by default
2	**Data access requests** – Access to data is handled restrictively and only authorized users with specific authorizations are granted access
3	**Baseline** – Data access is limited to the necessary minimum

⚠ Key Challenges

Hindrance to collaboration: Strict limitations on data access can make it difficult for teams to collaborate effectively

Knowledge silos: By restricting access to data, the need-to-know principle can unintentionally create knowledge silos within an organization

Limited innovation: Employees who don't have access to a wider range of data may be limited in their ability to identify innovative solutions

Potential delays: The process of requesting and granting access can lead to delays, especially for time-sensitive projects or situations where quick decision-making is crucial

Proud-to-share

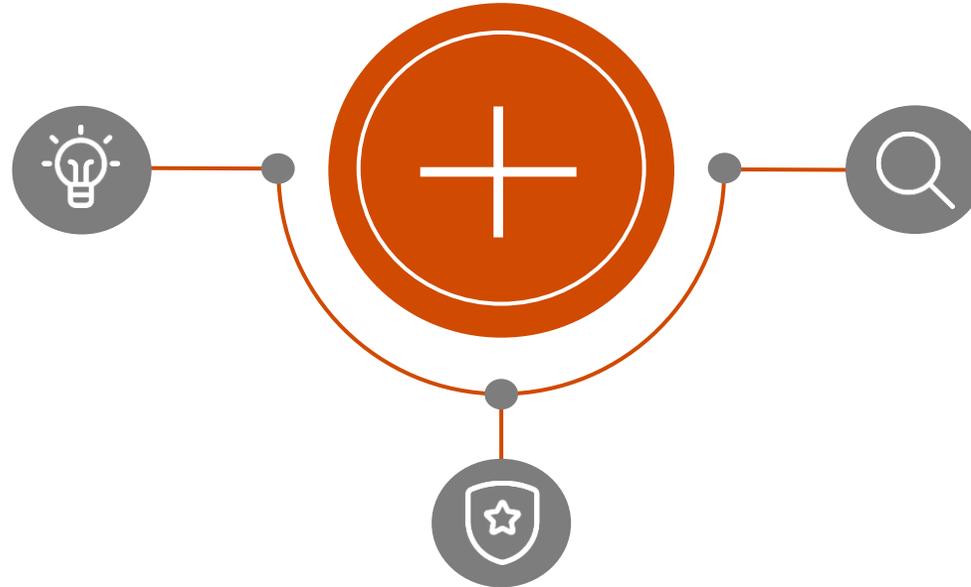|	
1	**Starting point** – All data is accessible by default, unless it is explicitly marked as deserving protection
2	**Data access requests** – Data access requests are only required for data marked as deserving protection
3	**Baseline** – Data access is not limited, unless there are specific restrictions

⚠ Key Impacts

Enhanced collaboration and innovation: Easier access to data fosters collaboration across teams and departments

Increased user productivity: When users can easily access the information they need, they can spend less time searching for data and more time completing their task

Improved use cases: By having access to more data, employees can make use-cases decisions that are more likely to be successful

Reduced costs: Streamlining data access processes and minimizing access restrictions can save organizations money on IT security resources

We have identified the following pain points during current data consumption

Transparency

There is **no central overview** of the **data** at SIGNAL IDUNA - **Each data consumer** has to build up his own **knowledge base** about the **available data**

Not all responsibilities are documented in a central repository. Any changes to data structures are communicated **to all data consumer individually**

Governance

Compliance with data retention policies and other **regulatory requirements** is ensured individually in a decentralized manner

Data quality is currently **monitored and assured** for all data consumers individually – in absence of central processes and guidelines

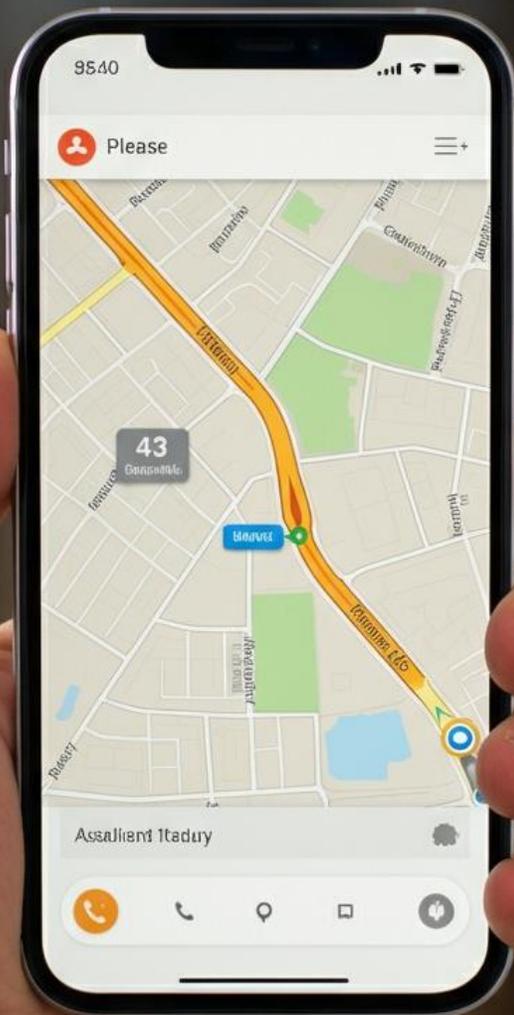Granting permission for data processing can carry significant **business risks**

High manual labor costs

For **individual** consumers **custom solutions** are developed and maintained

Numerous distinct steps are necessary to grant permission for data consumption, and each step requires **individual** identification

Currently, there is no standardized data shopping process

Where we want to go

Establish transparency regarding SIGNAL IDUNA's data diversity and create opportunities for effective data utilization.

Self-Service for data product provisioning
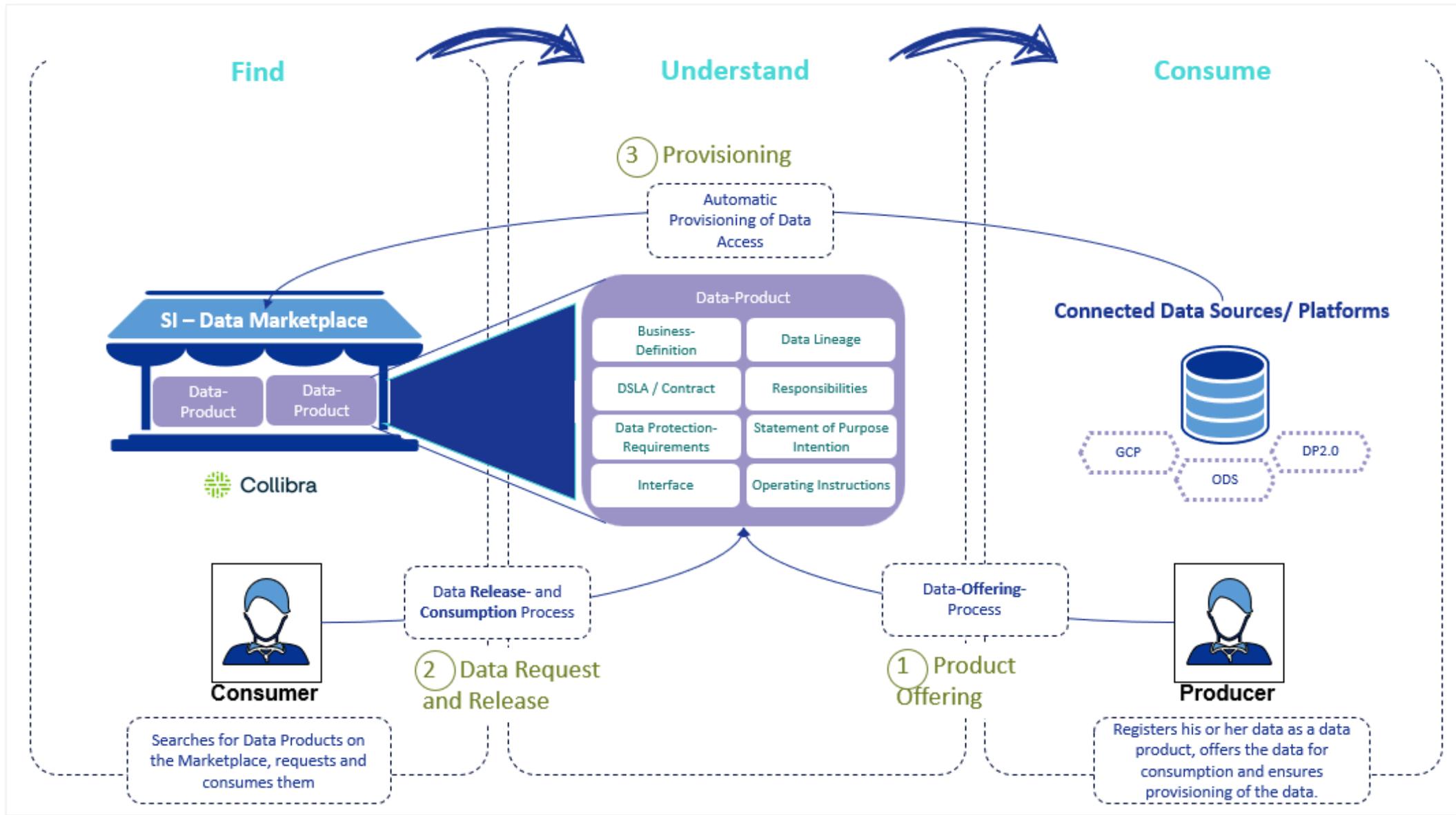(Data producers should be enabled to autonomously offer data products on the data marketplace in Collibra).

Self-Service for data consumers
(Users can independently understand and request data products based on the metadata).

Standardized and automated process for granting permission to access data products with forthright responsibilities.

Clear regulatory terms and conditions for data consumption and a central governance of the data usages at SIGNAL IDUNA.
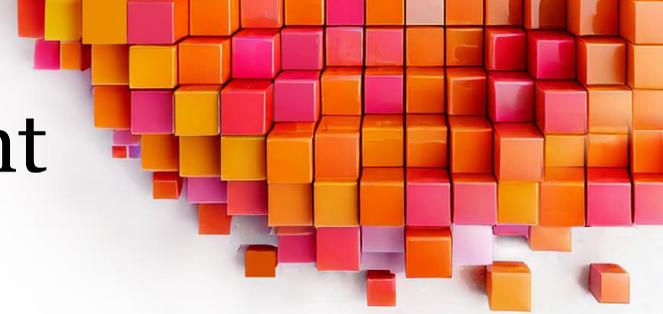
Find **Understand** **Consume**

③ Provisioning

Automatic Provisioning of Data Access

SI – Data Marketplace

Data-Product | Data-Product

Collibra

Data-Product

Business-Definition	Data Lineage
DSLA / Contract	Responsibilities
Data Protection-Requirements	Statement of Purpose Intention
Interface	Operating Instructions

Connected Data Sources/ Platforms

GCP   ODS   DP2.0

Consumer

Data **Release**- and **Consumption** Process

② Data Request and Release

Data-**Offering**-Process

① Product Offering

Producer

Searches for Data Products on the Marketplace, requests and consumes them

Registers his or her data as a data product, offers the data for consumption and ensures provisioning of the data.

02

How do we leverage GenAI for data management?

GenAI and data management are interdependent and drive each other forward

Data Management enables GenAI

GenAI **requires high-quality data management** to unlock its full potential, as its effectiveness depends on accurate data

- Inconsistent or poor-quality data undermines model accuracy and reliability

- Lack of efficient data pipelines leads to slower implementation and limited scalability

- Missed opportunities to maximize the value of AI-driven solutions due to data silos and inefficiencies

GenAI enables Data Management

Data management **struggles to achieve the efficiency** and quality required to handle complex, large-scale data challenges

- Reliance on manual processes leads to inefficiencies and delays in handling large-scale datasets

- Difficulty in extracting meaningful insights from unstructured and semi-structured data

- Ineffective integration and alignment of data across silos hinder decision-making

How can GenAI and Data Management profit from each other?

GenAI supports data management across multiple dimensions

Data Privacy & Security

- Labelling of sensitive data
- Dynamic semantic access authorization & control
- Identification of AI-generated (fake) data

Data Processing

- Generation of meta data & documentation
- Data mapping (source/target, technical/functional)
- Generation of test data and automated test execution

Data Analytics

- Self-service analytics & visualization
- Exploratory data analysis with generated queries
- Detection of data drift

Data Governance

- Detection of data lineage & data utilization
- Mapping of data objects and meta data objects
- Automated proposal for data ownership

Data Architecture

- Generation of architecture-compliant infrastructure as code (e.g. "Azure Console Copilot")
- Interactive guideline & counseling

Data Quality

- Generation of DQ rules including source code
- Suggestion & execution of data corrections
- DQ measurement without rules

With the help of GenAI, data fields can be efficiently assigned to the appropriate data domain

Situation

Compilation

Solution

**Definition of domain/s
Need of assignment**

Complex manual process

GenAI-based mapping based on semantic information from the data catalogue

Data Assignment

- 15 Data Domains have been defined to structure metadata of company
- Each domain is owned from business point of view
- All data must be assigned to at least one Data Domain, ideally on attribute level, to ensure clear governance and responsibilities

- Manual assignment of data fields to domains is time-consuming
- Employees have to analyze each data field individually
- This process can be tedious with large data sets
- There is a risk of misclassification and incorrect categorization.
- Subjective judgements and individual interpretations can lead to errors

The solution is to use GenAI to map data fields to their respective domains. This process is facilitated by prior descriptions of the data domains so that manual and time-consuming efforts are no longer required.

We have built this MVP mapping data fields to the correct data domain using GenAI

Goal

The aim is to assign data fields (data attributes/data entities) to the corresponding data domains, which are previously described by describing the data domains.

Functionality

GenAI (Vertex AI) uses metadata to automatically assign data fields to the corresponding data domains. The process begins with the provision of the data domains and searched data fields. Vertex AI then analyses patterns in the data to perform automated categorization.

Method

- Input: Data domains (metadata) description and explanations, Data to be assigned
- Output: Assignment (data attribute: associated data category) and relationship in Collibra

Requirements

- Metadata, data understanding, access to ML/AI, AI knowledge, data modelling, prompt engineering

Data Capabilities

Data Strategy → Data Architecture → Data Quality → Data Governance → Data Processing → Data Protection → Data Analytics

Chances

- Improved data management: Precise assignment of data fields to suitable data domains improves data quality
- Increased efficiency: Faster, simplified assignment processes increase efficiency and reduce errors
- Consistency and compliance: Automatic assignment strengthens compliance and minimizes data protection risks

Risks

- Incorrect assignment: If the data domains have been described incorrectly or the data fields can be assigned to several data domains, there is a possibility of incorrect assignment

Market readiness

Concepts
Analysis of requirements and ideation
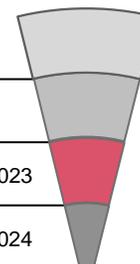
Proof of Concept
Practical proof of concept

MVP
First functional version — 2023

Adoption
Implementation — 2024

GenAI Impact

GenAI is used for the precise assignment of meta information.

Exemplary in- and outputs of the use case showcase its potential to facilitate the former manual assignment

Input

- Data domains with description
- Explanation of data domains (metadata)
- Data to be assigned

Example

> ⌄ Prompt   🗑 CLEAR PROMPT
>
> Im Folgenden findest du unsere Definition einer Datenkategorie und im Anschluss die Datenkategorien unseres Modells.
> Eine Datenkategorie ist eine Klassifizierung von Daten, die auf gemeinsamen Merkmalen oder Eigenschaften basiert. Datenkategorien werden verwendet, um Daten zu organisieren und zu strukturieren, und um den Zugriff auf Daten zu erleichtern. Es gibt viele verschiedene Arten von Datenkategorien, z. B. personenbezogene Daten, Finanzdaten, medizinische Daten, Kundendaten, Produktdaten usw.
> 1. Einkauf: Einkauf bezieht sich auf die Daten, die mit dem Einkauf von Gütern und Dienstleistungen in Verbindung stehen. Dies schließt Lieferanten, Kaufverträge, Konditionen und die Leistungsbeschreibungen mit ein. Außerdem ist die Pflege und Katalogisierung von bereits bezogenen oder beziehbaren Produkten ein Teil der Kategorie. Preisliste, Laufzeit Information, Inventarliste, Leistungsbeschreibung
> 2. Finanzen: Finanzen beinhaltet alle Daten zu finanziellen Aspekten der .... Dazu gehören Buchhaltung, Budgetierung, Vermögensverwaltung und Investitions Informationen. Ein weiterer Teil der Kategorie umfasst die finanzielle Berichterstattung gemäß Solvency II und der Bilanzierung.
> <...>
> Meine Frage: Welche der folgenden Datenkategorien gehört zu welcher Datenkategorie, die oben genannt wurden?
> Strategische Planung und Krisenmanagement, Talent Acquisition und Integration, Tarif, Umwelt Information, Underwriting und Bewertung, Unternehmenskommunikation, Ursache und Sachverhalt, Verarbeitung Tätigkeit, Versichertes Objekt
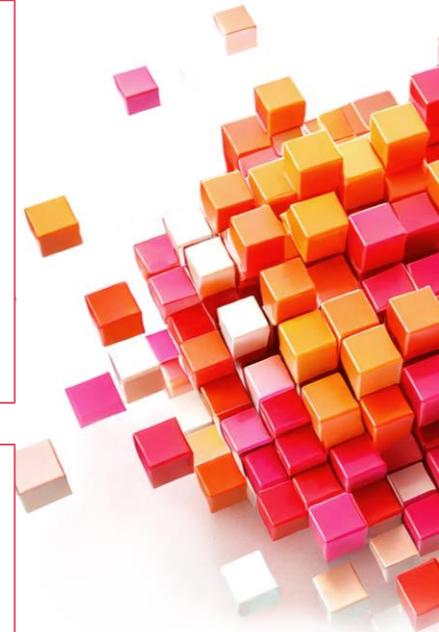>
> **SUBMIT**

Output

- Assignment (data attribute: associated data domain) and relationship in Collibra

Examples

> ⌄ Response
>
> - Strategische Planung: Strategie und Steuerung
> - Krisenmanagement: Organisation
> - Talent Acquisition und Integration: Personal
> - Tarif: Produkt
> - Umwelt Information: Markt
> - Underwriting und Bewertung: Rückversicherung
> - Unternehmenskommunikation: Organisation
> - Ursache und Sachverhalt: Schaden und Leistung
> - Verarbeitungstätigkeit: IT
> - Versichertes Objekt: Versicherungsvertrag

> ⌄ Response
>
> Schadenbericht: Schaden und Leistung
> Leistungsabwicklungsbericht: Schaden und Leistung
> Schadenspräventionsbericht: Schaden und Leistung
> Anspruchshöhe: Rückversicherung
> Berichterstattung: Rückversicherung
> Kondition: Rückversicherung
> Zeiterfassung: Personal
> Mitarbeiterzufriedenheit: Personal
> Steuer-ID: Person
> Abteilungsstruktur: Organisation
> Wettbewerbsanalyse: Markt

> ⌄ Response
>
> Standortverwaltung: Organisation
> Leistungs Abwicklungsbericht: Schaden und Leistung
> Produktlebenszyklus: Produkt
> Neuentwicklung: Produkt
> Umweltbericht: Markt
> Wissenschaftliche Herausgabe: Markt
> Laufzeit Information: Einkauf
> Steuer-ID: Person
> Physische Assets: Finanzen

GenAI was leveraged to identify root causes and measures against DQ incidents

Key questions

- Are there any trends in the data? Do certain incidents occur in a pattern?

- Which solutions have led to the incidents' resolution? Can we identify a pattern and leverage this for the future?

- Can we identify and categorize root causes for certain incidents?

- How can we identify duplicates or wrongfully opened incidents?

Situation

➤ Client was facing approximately 500,000 ServiceNow incident records along with multiple meta tables

➤ Client was asked to extract actionable insights and demonstrate how generative AI can be leveraged in addressing previously unresolved issues

Approach

➤ Deployed a combination of cloud and local LLMs based on GPT-4o and Llama 3.1

➤ Designed complex analysis chain including incident categorization into buckets, root cause identification, trend identification, etc.

➤ Processed 500,000 ServiceNow incidents

Results

➤ Identified >80 incident buckets and refined them to 50 actionable buckets during client workshops

➤ Determined root causes for each bucket by employing LLMs and metadata

➤ Developed a comprehensive final report highlighting key findings and actionable tasks for each specific bucket
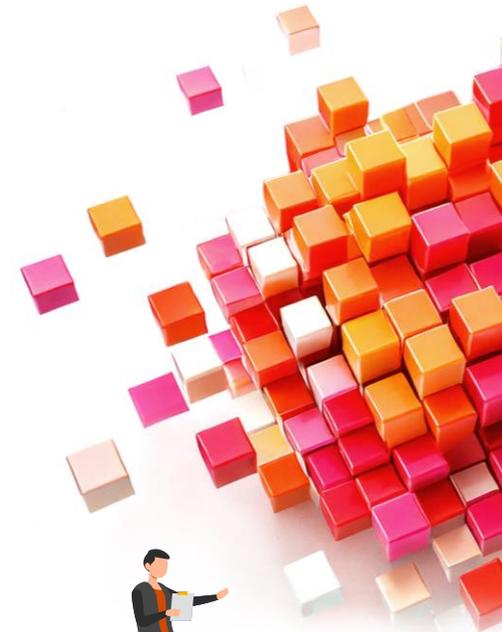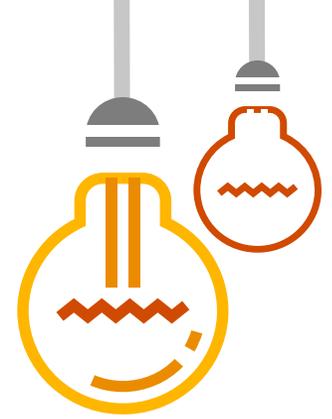
Let's discuss...

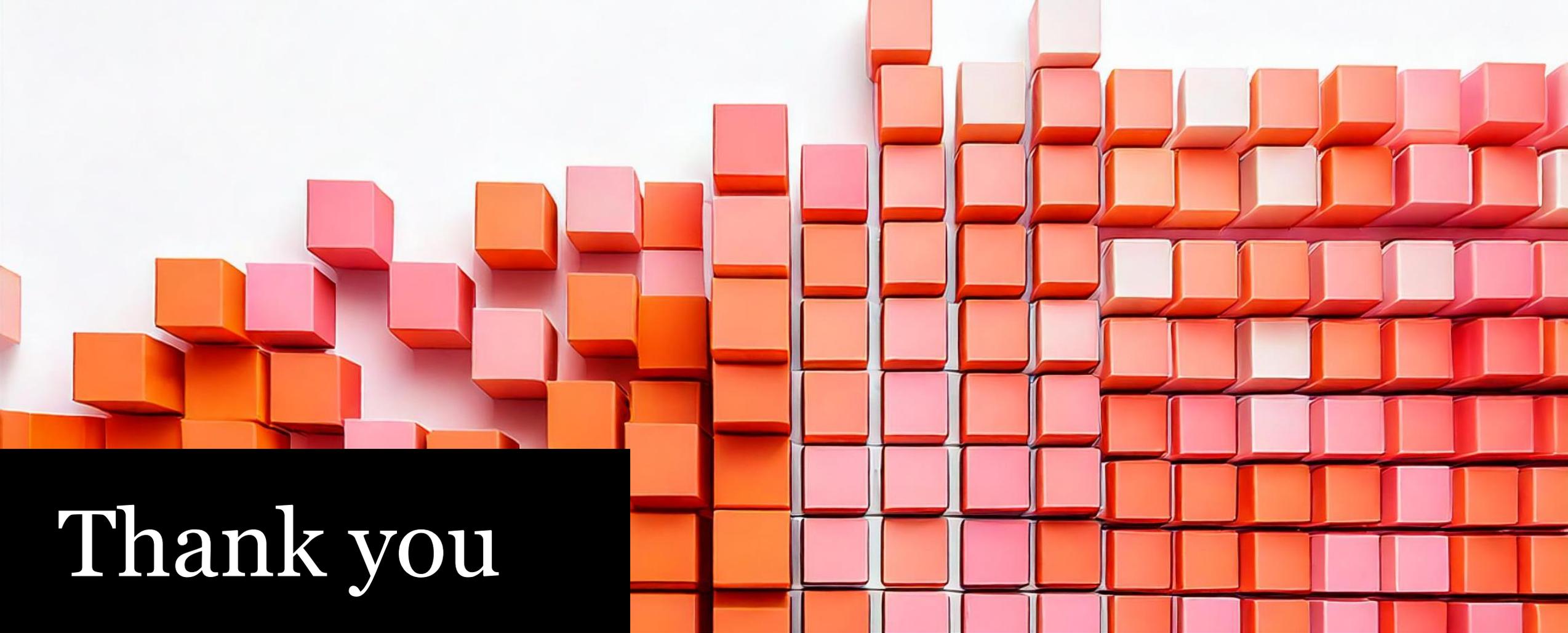What are your current challenges with operationalizing Data Management?

Have you adopted a data domain and data product approach? If so, how, to what success?

How do you leverage GenAI? Which are the most promising areas to your firm?

Thank you