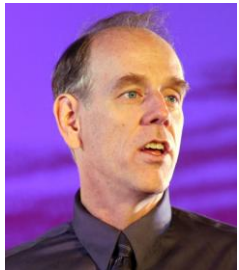


# EDM Webinar

## Data Products: Emerging Uses and Generative AI for Deployment

*A conversation with*



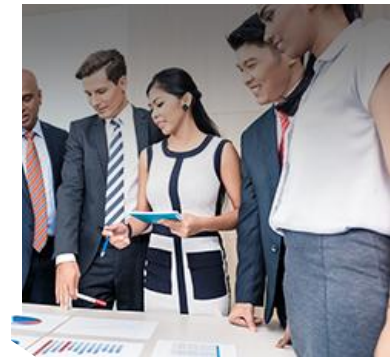
**Stephen Brobst**  
CTO  
Ab Initio  
Software



**Alessandro Allini**  
CDO  
Crédit Agricole  
Italia



**Rob Parks**  
Senior Engineer  
Ab Initio  
Software



# Today's panel

Moderator



**John Bottega**

President  
EDM Council



**Stephen Brobst**

Chief Technology Officer  
Ab Initio Software



**Alessandro Allini**

Chief Data Officer  
Crédit Agricole Italia



**Rob Parks**

Senior Engineer  
Ab Initio Software





## Seven characteristics of the **Data-driven** enterprise:

- 1** Data is embedded in every decision, interaction, and process
- 2** Data is processed and delivered in real time
- 3** Flexible data stores enable integrated, ready-to-use data
- 4** Data operating model treats data like a product
- 5** The Chief Data Officer's role is expanded to generate value
- 6** Data-ecosystem memberships are the norm
- 7** Data management is prioritized and automated for privacy, security, and resiliency

Excerpted “*The data-driven enterprise of 2025*” (McKinsey).

# Revised Direction: Data as a Product

## Traditional Thinking

- Data as "an asset"
- Subject-area organization of data
- Source-driven
- Store it (all) in a warehouse
- Measure of success is terabytes stored
- Build it and they will come
- Focus on descriptive analytics and reporting
- Centralized BICCs, DICCs
- Metadata (IT-driven)

## New Direction

- Data as "a product"
- Business process domain organization of data ownership
- Usage-driven
- Curated for consumption
- Measure of success is the value created
- Advanced analytics with prediction and prescription
- Emphasis on product life-cycle
  - Prototype, design, test, integrate, deploy
  - DataOps / Agile methodologies
  - Business product owners
- Observability of consumption and value
- Metadata (business-driven)

**Data as a Product:** *Packaging of data that is consumable by virtue of being discoverable, understandable, curated, having self-describing semantics, trustworthy, usage-driven, re-usable and interoperable.*

# The Future of Data

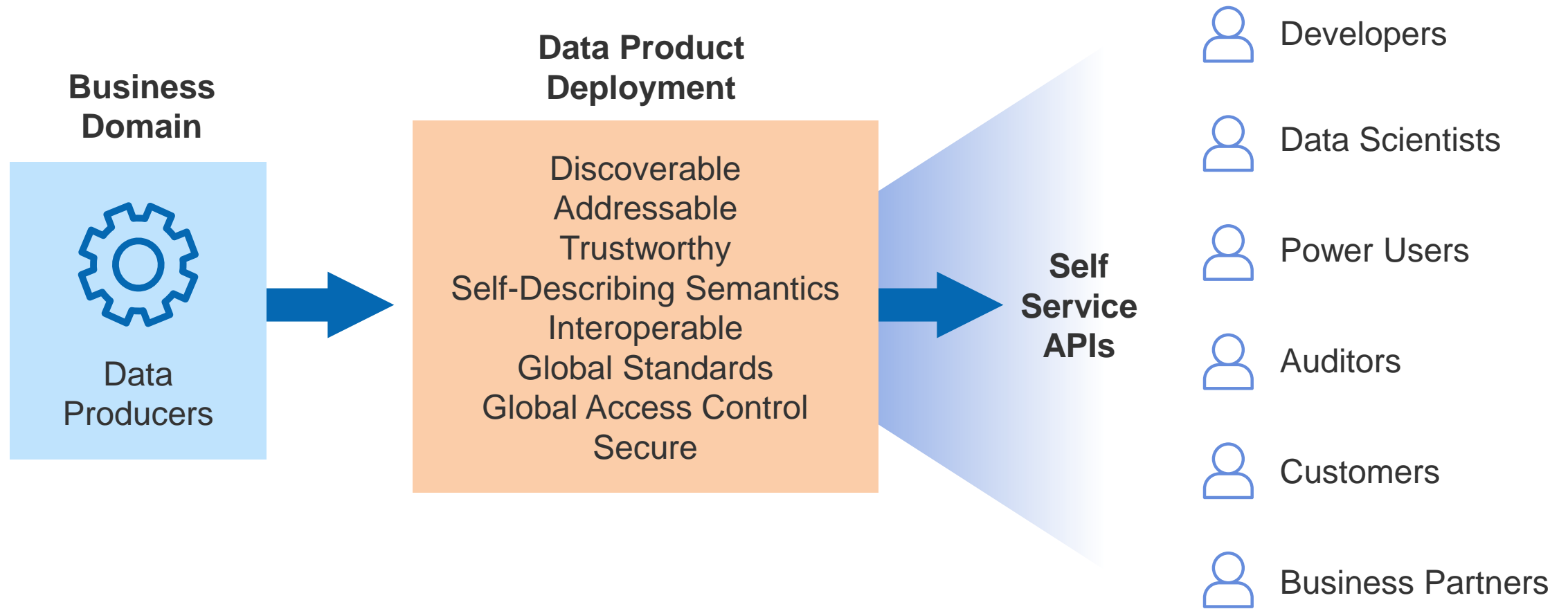
“

If you want to be data-driven...

**...you need to be metadata-driven.**”

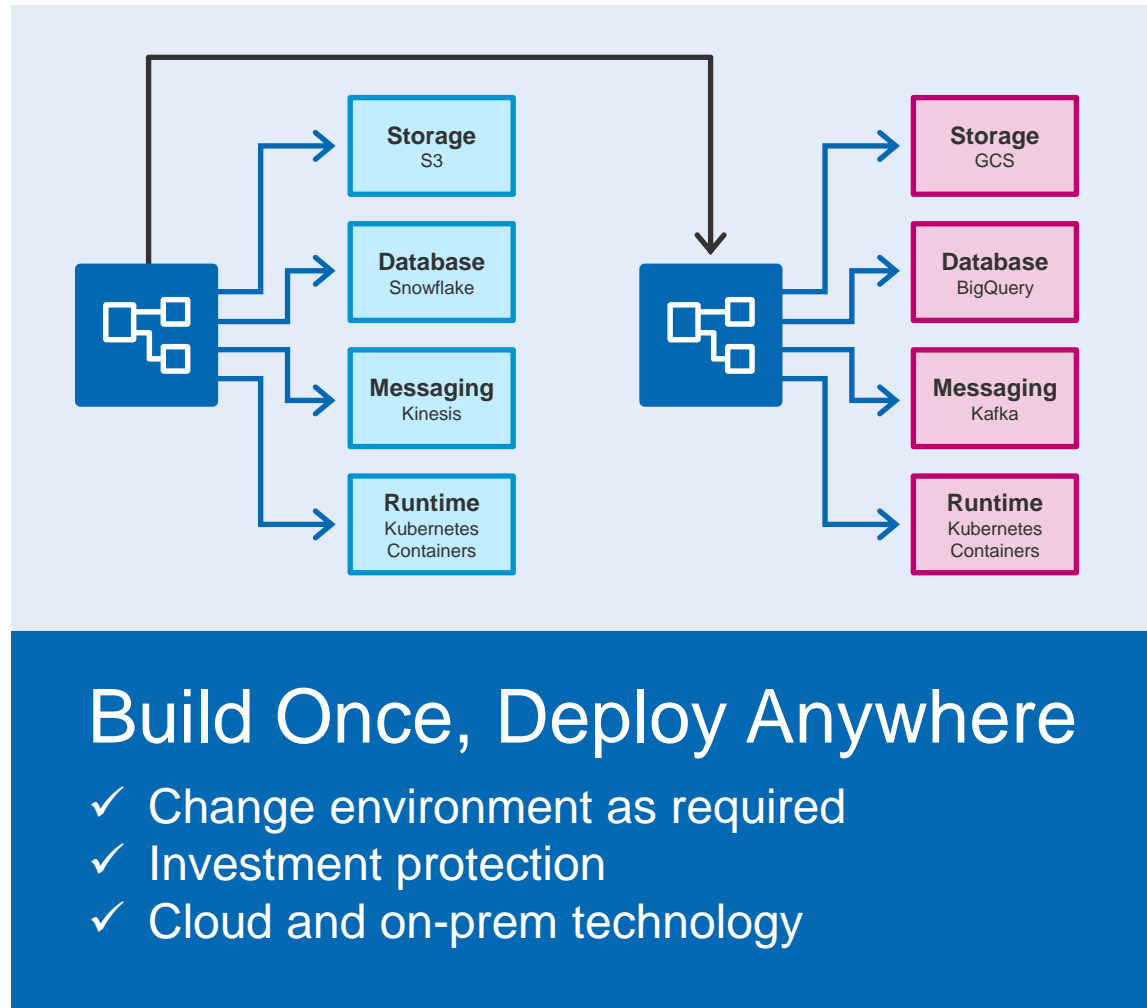
–Gartner Data and Analytics Summit, London, 2024

# Anatomy of a Data Product



# Architectural Principles:

## Metadata-Driven / Just-In-Time Execution



### Cloud Storage

- Amazon S3
- ADLSg2
- Google Cloud
- Amazon EFS, EBS
- HDFS

### Cloud Messaging

- Amazon Kinesis
- Amazon MQ
- Kafka (MKS)
- Google Pub Sub

### Cloud Database

- Google BigQuery
- Amazon Redshift
- Azure SQL Server
- Snowflake
- Teradata

### On-Premises Data

- Oracle
- z/OS, AIX, DB2/400
- VSAM/ Mainframe files
- IMS Mainframe
- HDFS

### On-Premises Compute

- z/OS Mainframe
- Linux/ Unix
- Windows
- Hadoop

### Runtime Environments

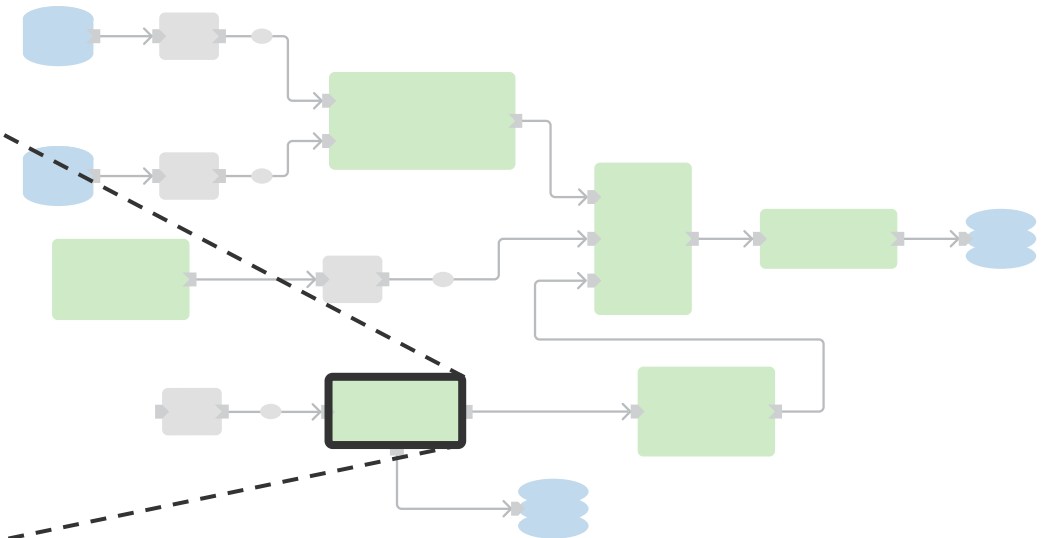
- Containers
- Kubernetes
- Hadoop
- Google Dataproc
- Amazon EMR
- Virtual Machines

# Architectural Principles:

## Low / No Code with Self Service Computing

- Co-pilots using Generative AI for constructing data products
- Co-pilots using Generative AI for conversations with data as well as metadata
- Build and deploy integration patterns
- Create and manage metadata
- Empower business users

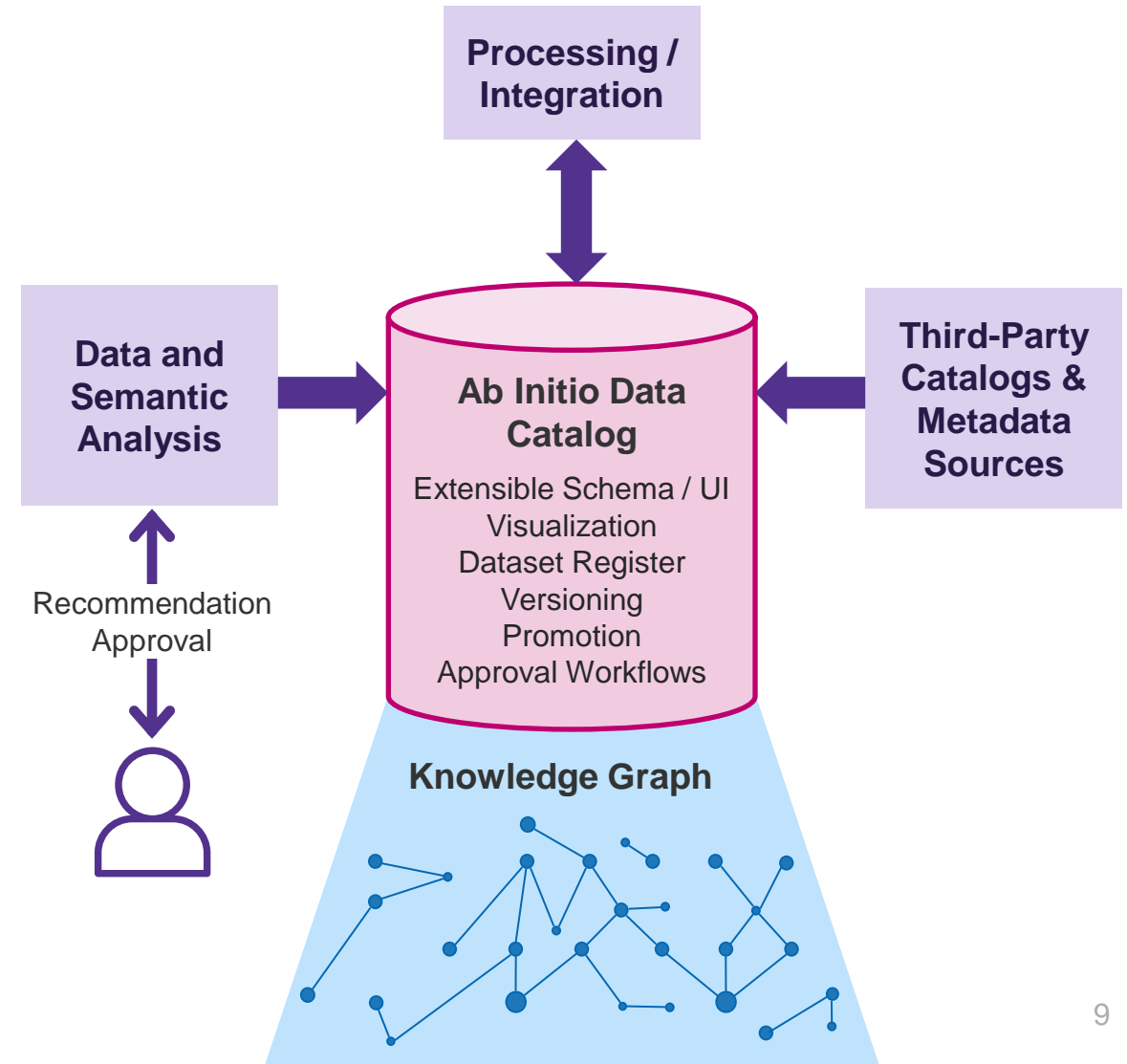
Triggers					Outputs
Avg Monthly Balance	Avg Monthly Changes	Credit Score	Years as Customer	is true:	Award Level
>=20000	>=1000	excellent	>=10	any	Titanium
>=19000	>=1000	↓	↓	any	Gold
>=2000	>=5000	excellent or good	>=2	any	Silver
>=200	>=500	poor	↓	Total Entertainment >10	Bronze
any	any	any	any	any	No Reward



# Architectural Principles: Active Metadata

Metadata is core to all processing and governance activities:

- Understanding and sharing
- Data quality, masking, lineage, & controls
- Reference data management
- Driving integration processes
- Context for optimizing Generative AI using LLMs





Ab Initio

# Data Product Partnership – Crédit Agricole Italia & Ab Initio

EDM Council, 02/27/2025

# PROGRAM OBJECTIVES

## Bridging the Gap Between Business and IT

THE GOAL IS TO FACILITATE **INTERACTIONS BETWEEN BUSINESS AND IT**, CREATING REAL VALUE FOR THE BANK  
**THROUGH THE ADOPTION OF A DATA-CENTRIC APPROACH**

DATA



IT

ACTION



BUSINESS

DATA DRIVEN BANK

### GOALS



#### QUALITY

Enhanced Data Quality  
(including its validation  
and centralization of  
controls)



#### OPERATIONAL EFFICIENCY

Operational (e.g. *time-to-market*) and cross-directional  
process efficiency



#### DATA GOVERNANCE

Compliance with both  
international and local data  
governance regulations

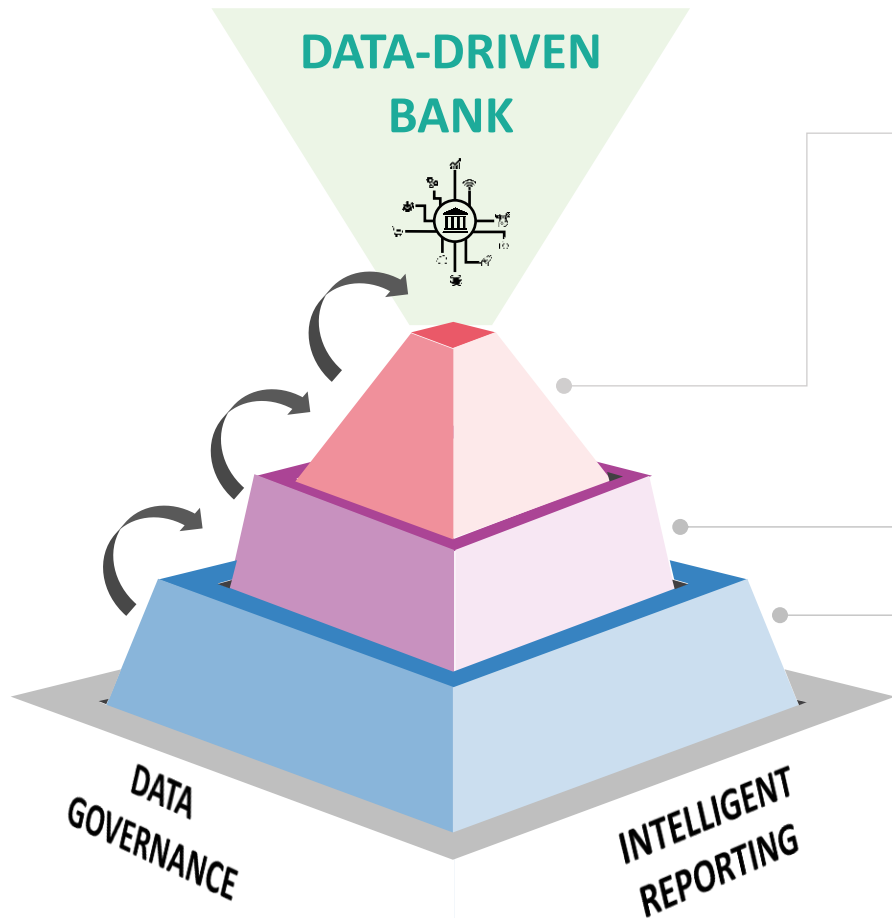


#### OPERATIONAL MODEL

Evolution of the operational  
model to enable innovative  
advanced analytics use-cases

# THE INITIATIVE

## The Path to a Truly Data-Driven Bank



THE MODEL FOR A **DATA PLATFORM** CAN CONSIST OF **3 PHASES** THAT PROGRESSIVELY ENABLE **FULL KNOWLEDGE AND USABILITY OF INFORMATION ASSETS**

### ANALYTICS LAB

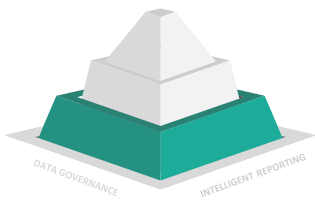
- Introduction of **advanced analytics tools** to support decision-making processes of IT and business functions
- Review of the **data management organizational model**, with the aim of reinforcing analytics elements and establishing a connected hub
- Use of **analytics to enhance Data Quality and Data Governance**

### NEW MODELS & PROCESSING ENGINES

- **Data enrichment** through specific processing engines for each area (e.g. *Planning & Control, CRM, Risk, Compliance, etc.*), gathered into a comprehensive Data Model
- **Convergence** of the Bank's areas and availability of **cross-areas indicators**

### DATA CENTRICITY

- Creation of a **new Data Platform** feeding through:
  - Integration of the information perimeter with **internal data** (e.g. *CRM, Risk, Compliance, etc.*)
  - Enrichment with new **external data**
- **Rationalization** of existing systems



OUR TRANSFORMATION APPROACH WAS DRIVEN BY **3 KEY ELEMENTS**, WITH THE AIM OF CREATING A DATA PLATFORM ABLE TO RESPOND TO **FUTURE NEEDS**

### 1 Future-proof infrastructure

Support agility, availability, protection, and performance requirements based with an **on-demand capacity model** that allows for hybrid & multi-cloud deployments

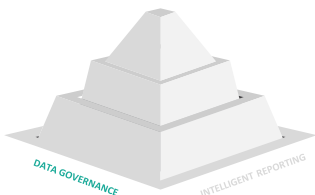


### 2 Low-code/no-code approach

Automation during data preparation (ingestion, curation and provisioning) to **reduce costs and time-to-market for analytics solutions**. Reduce the need for advanced programming skills in favor of **metadata-driven modules**

### 3 Native Data Governance

Enable Data Scientists, Engineers and Business Users to **analyze data through search-based UIs** based on **business metadata** with complete mapping to technical metadata



CRÉDIT AGRICOLE ITALY HAS ADOPTED A **DATA GOVERNANCE EVOLUTIONARY PATH** WITH THE AIM OF REPLACING THE “**CLASSICAL**” **APPROACH** CHARACTERIZED BY **HIGHLY MANUAL PROCESSES**

THE ADOPTION OF **AB INITIO** ACHIEVES THE **FOLLOWING GOALS**:



### DATA GOVERNANCE BY DESIGN

Automated data/metadata mappings keep information up-to-date and follows the evolution of information systems



### ENHANCEMENT OF THE DEVELOPMENT PROCESS

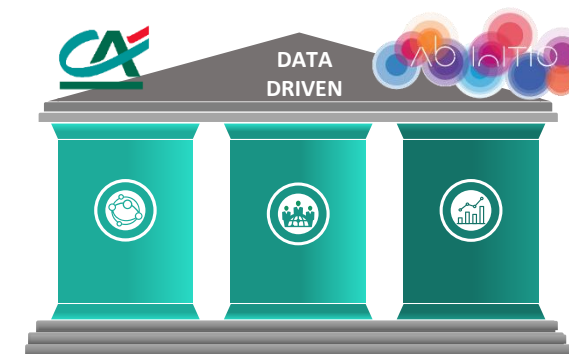
Automation and AI-driven linkages between logical and physical assets minimizes involvement of Data Owners and allows for the creation of Data Products



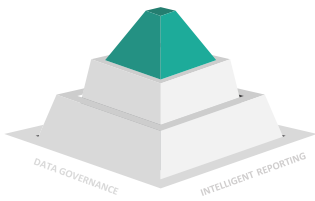
### ACTIONABLE DATA GOVERNANCE

Comprehensive Data Lineage and Knowledge Graph technology enables the Bank to launch sophisticated controls and analytics

### ADVANCED APPROACH

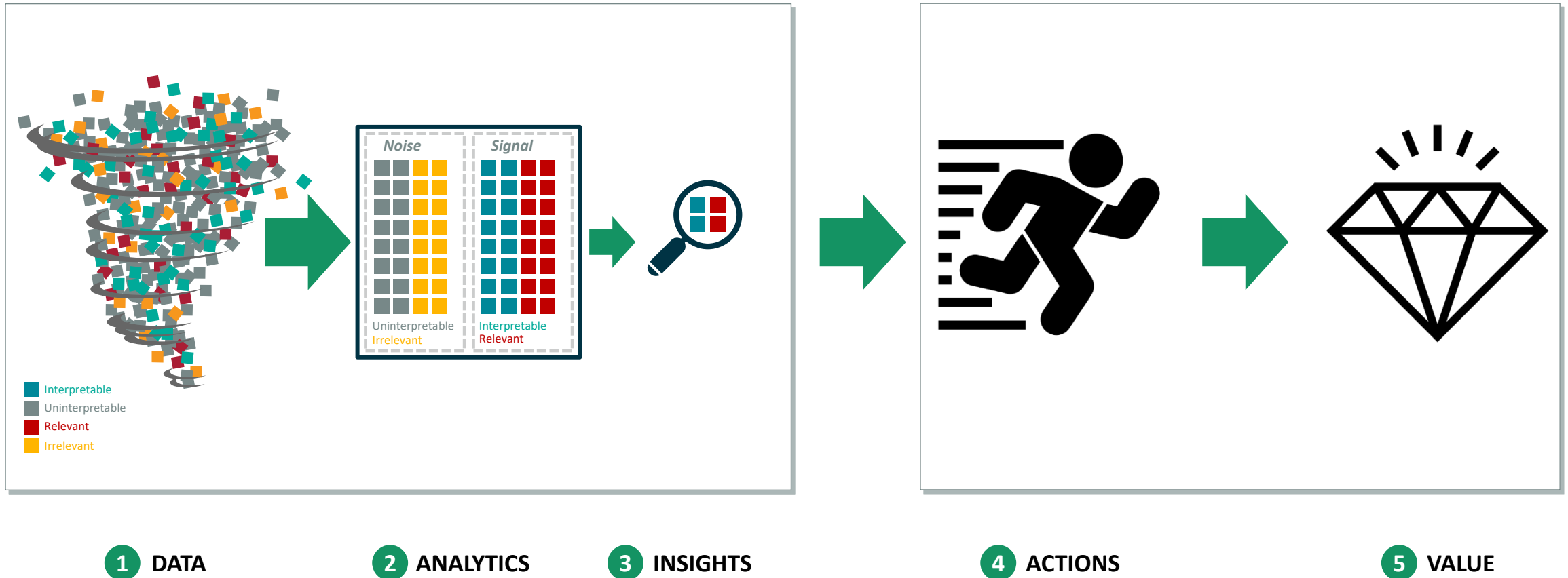


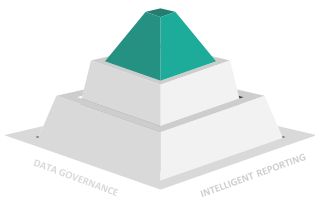
DATA GOVERNANCE



From data to insight – From insight to action

IN THE DIGITAL ERA, **ADVANCED ANALYTICS** ARE ESSENTIAL TO EXPLOITING THE TRUE VALUE OF DATA

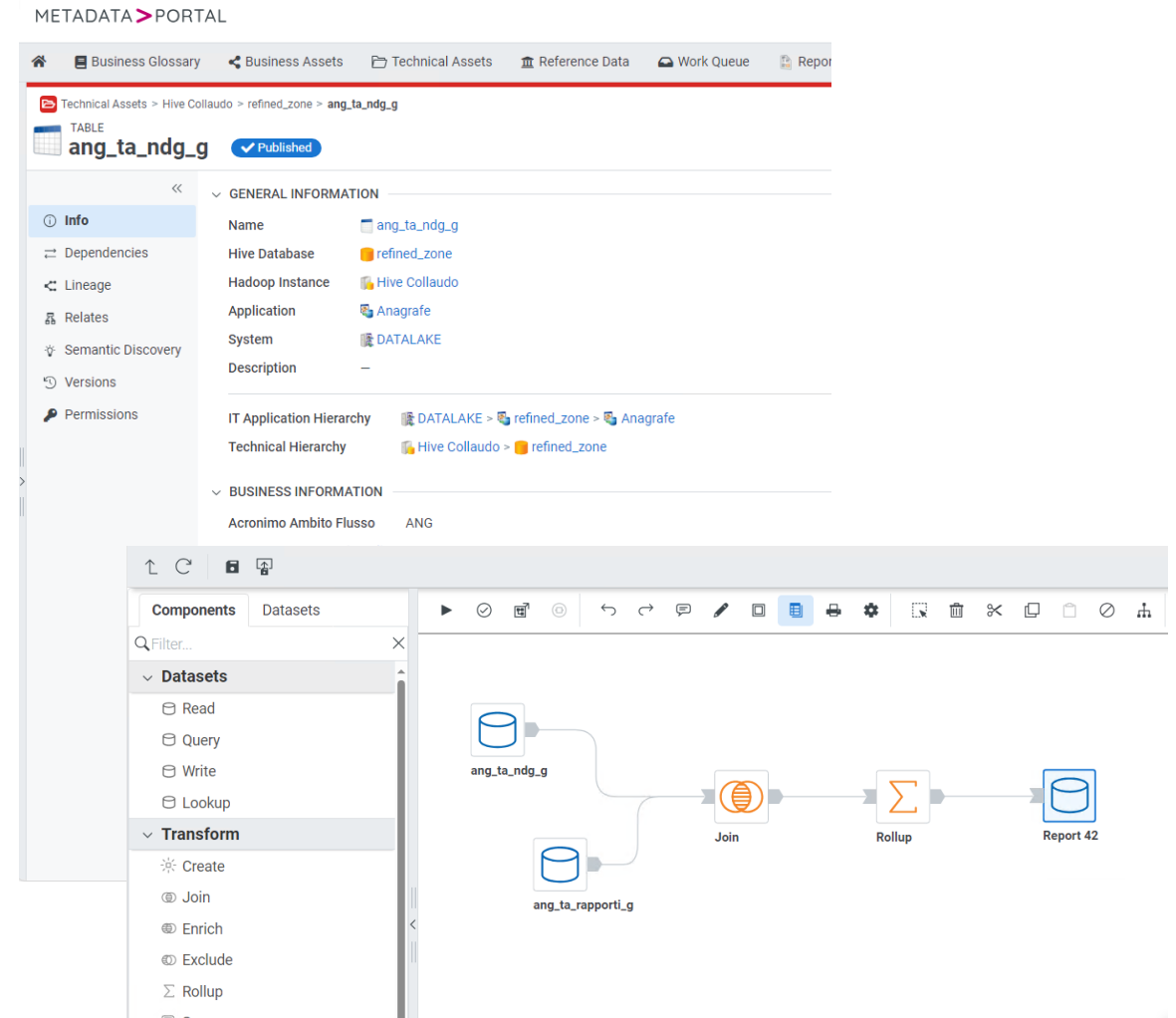


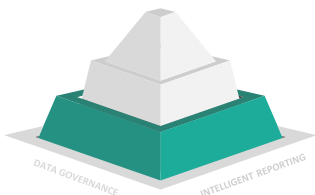


# Data Product Approach

## Data Products used by Business Units

- **Data Products** come from **advanced data pipelines** with:
  - Data **Governance**
  - Data **Quality & Controls**
  - Data **Masking & Cleansing**
- **Data Products** follow a **proper naming convention** and are grouped by business domains into 2 zones:
  - Refined zone with masked data
  - Restricted zone with unmasked data
- **Datasets** exposed are linked with **metadata** and provide a description for **each data element**
- Each business unit has a **dedicated project sandbox** and use Ab Initio Easy>Data to **produce new datasets** that are **registered and governed**.



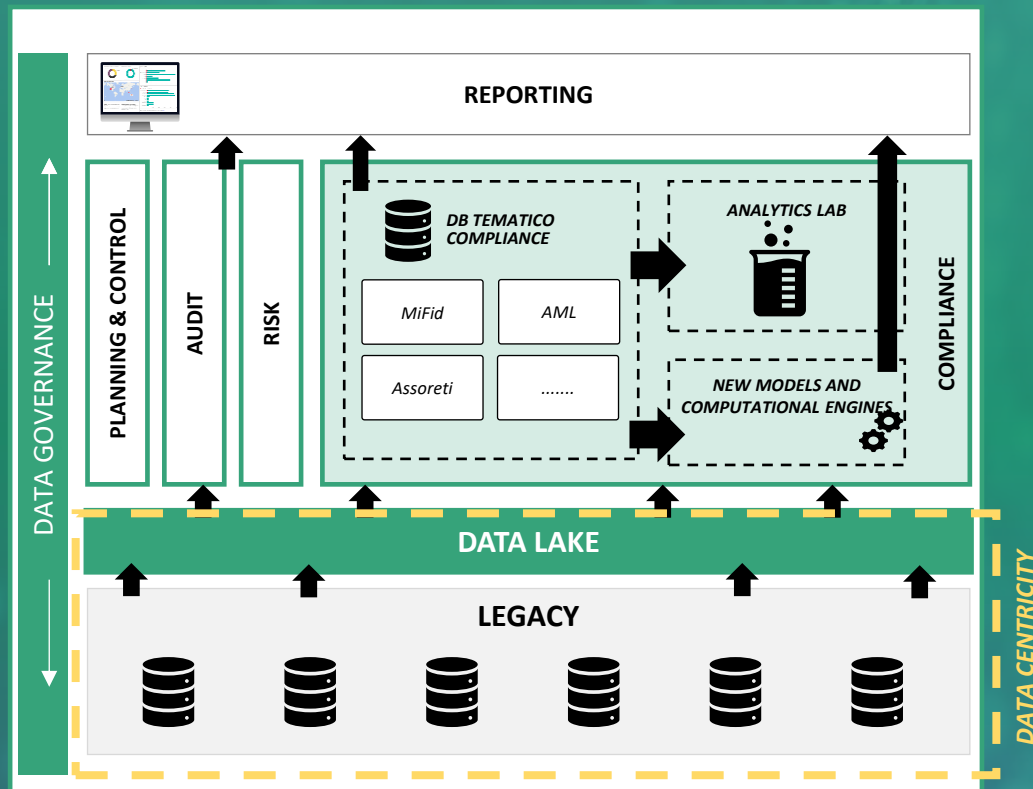


# DATA CENTRICITY

## The Data Platform

### "DATA PLATFORM"

The **DATA PLATFORM** enables the evolution from an **operational model** based on **static and manual processes** to an **integrated model** characterised by high **automation, external integration** and development of **advanced logics for data management**



- **Full Data Lineage** at all layers and technologies
  - Legacy
  - Data Lake
  - Analytics Lab, etc.
- Data Access to Business Units
  - **Virtualization** of Data Stores
  - **Data Product** and Data Marketplace
- Working with Ab Initio
  - Strong Collaboration
  - High Competency

# Core Components for Delivering Data as a Product

- Blueprints with metadata-driven automation for data product manufacturing using Generative AI
- Business and technical metadata
- API framework
- Data marketplace
- Business catalog with business definitions and semantic modeling
- Self-service access to data via low/no code environments using LLMs
- Lineage, transparency, and auditability for all data
- Data quality service level management with self-healing data
- Data freshness service level management with streaming, micro-batch, and batch processing
- AI-enabled with Generative AI as well as traditional AI/ML algorithms.

# Fastest Path to Data Products

- Use of Generative AI powered by an LLM and underlying metadata to automate creation of data products
- Comprehensive data governance to create trust in the data product content with lineage, data quality rules, access contracts, semantic layer aligned to business knowledge workers
- API framework for programmatic access to data as well as business self-service access to data via no/low code and Generative AI
- Data marketplace allowing business knowledge workers to shop for data using rich metadata enabled for discovery (by humans or AI algorithms)
- Data products can be virtualized or physicalized (or hybrid) using best practices Data Fabric capabilities.

# Call to Action

## Embrace Data as a Product

- Package data to be reused across multiple knowledge worker communities
- Align semantic models to knowledge worker communities
- Establish service levels for data (time, durability, quality)
- Deliver radical self-service
- Move with agility from discovery to production
- Determine a balance among conflicting goals or different perspectives

## Embrace Generative AI with Metadata Context

- To accelerate construction of data products
- To drive natural language conversations with data
- To drive natural language conversations with metadata

## Embrace the Ecosystem

- Support for heterogeneous platforms, as well as an evolving LLM landscape
- Leverage best of breed services and LLMs from both public and private clouds, including open source and ISVs

# Questions?

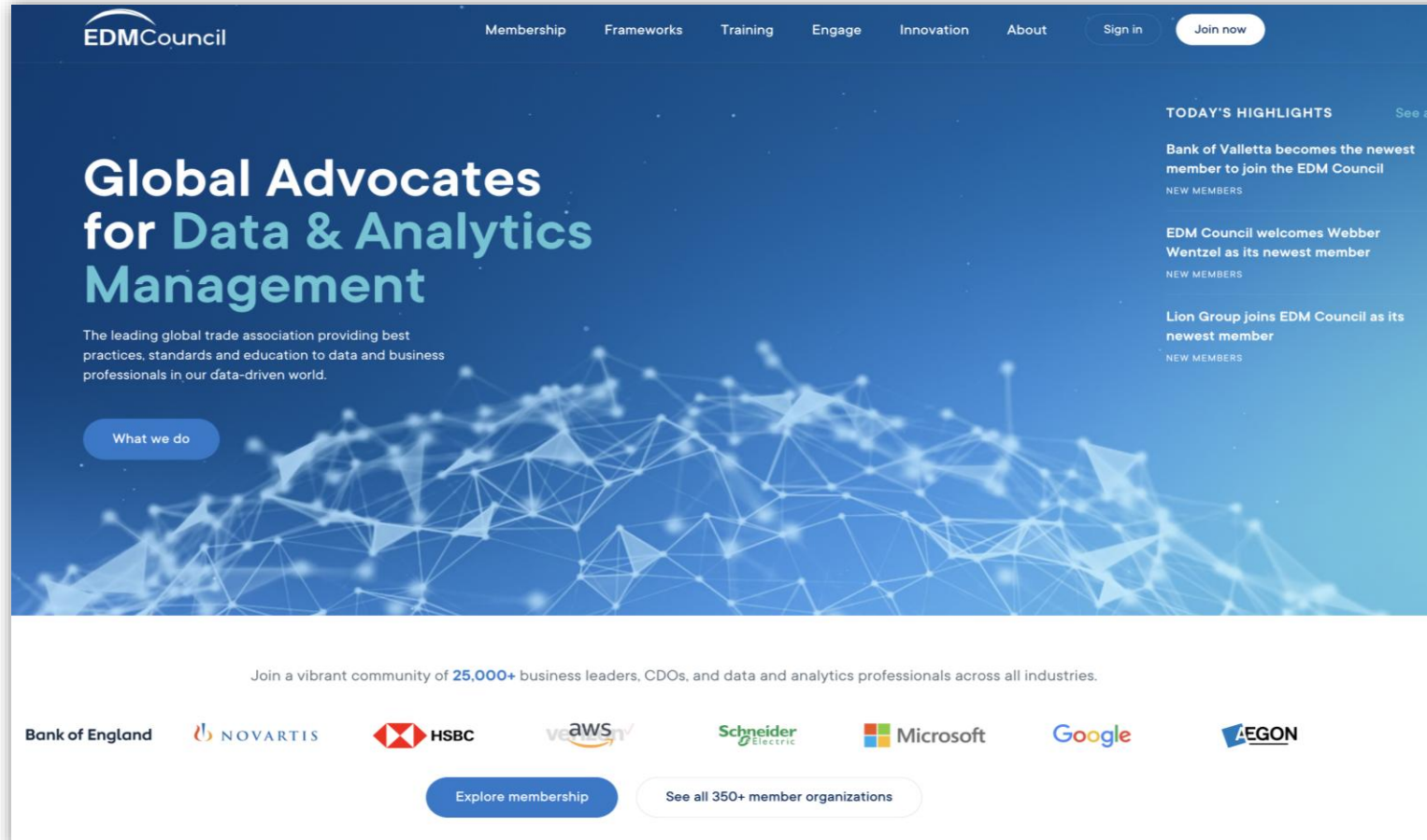


[solutions@abinitio.com](mailto:solutions@abinitio.com)

**EDM** Webinar 

---

# Join EDM Council and our membership community of companies...



**350+ Member Firms**

Cross-industry,  
including Regulators



**25,000+**

Professionals



**Worldwide**

Americas, Europe,  
Africa, Asia, Australia

**edmcouncil.org**



EDM Webinar 

Thank you!

**FOR MORE INFORMATION:**

Doug Housman  
Senior Account Manager  
Ab Initio Software  
*[dhousman@abinitio.com](mailto:dhousman@abinitio.com)*

