

# EDM Webinar

## AI, Data & Analytics Controls (ADAC): Using CDMC for AI Use Case Governance

*A conversation with*



**Oli Bage**  
Head of  
Architecture,  
Data & Analytics,  
LSEG



**Lexy Kassin**  
Lead Data & AI  
Strategist,  
Databricks



Moderator



**Jim Halcomb**

Head of Product Management  
EDM Council



**Oli Bage**

Head of Architecture, Data  
& Analytics,  
LSEG



**Lexy Kassan**

Lead Data & AI Strategist,  
Databricks



# Oli Bage

## Head of Architecture, Data & Analytics, LSEG

As Head of Architecture for London Stock Exchange Group's Data & Analytics business, Oli partners with customers, vendors and teams across sales, product, engineering, and operations to design the industry's next generation data, analytics and workflow platforms.

Oli founded the Cloud Data Management Capabilities (CDMC) initiative, an EDM Council initiative of 160 banks, cloud providers and emerging technology companies, which has released a framework of capabilities, best practices, and automated controls for handling sensitive datasets on cloud. In 2024 the group has developed additional capabilities for AI governance (ADAC) and data products. Oli also contributes to the OMG DPROD standard and FinOS CCC and Architecture-as-code projects.

Before joining LSEG in 2020, Oli worked for Morgan Stanley for 21 years. He was most recently the Distinguished Engineer in Morgan Stanley's Centre of Excellence for Data and chaired the investment bank's architecture committee.

Oli is a passionate supporter of the emerging technology ecosystem and regularly hosts the London Enterprise Technology Meetup in LSEG's offices. Oli is an angel investor in emerging data technology companies and holds a degree in Computer Science from the University of Cambridge.



# Lexy Kassan

## Lead Data & AI Strategist, Ethicist

Lexy Kassan is a Lead Data and AI Strategist at Databricks, overseeing and guiding customer organizations through holistic transformation programmes built on industry trends and best practices. As an Ethicist, Lexy coaches companies on implementing responsible AI and works with public affairs to shape Databricks' messaging to regulators and lobbyists.

Outside of work, Lexy mentors for multiple initiatives supporting underrepresented groups in STEM. She is also the Founder and Host of the Data Science Ethics Podcast in which she delves into the responsibilities and societal ramifications of applying AI across both private and public sectors.



# CDMC: Industry Engagement

100+ Leading firms and 300+ participants actively participating since May 2020




## CDMC Working Group



## Cloud & Technology Provider Certification



## Regulatory Engagement

- 
- **US:** Federal Reserve, CFTC, FDIC, SEC
  - **Canada:** OSFI
  - **UK:** BoE, FCA, ICO
  - **Germany:** BaFin
  - **Swiss:** FinMA
  - **Australia:** APRA
  - **Singapore:** MAS
  - **African Continental Cloud Regulatory Symposium**
  - Others

## EDMC Support

- 
- **Training Courses**
  - **Assessments & Certification**
  - **Authorized Partner Program**
  - **FINOS Open Source Project**
  - **Other industries engaged:** Pharma/Life Sciences, Telecommunications, Manufacturing, Energy, Insurance, Government

## CDMC+ Extensions

- 
- Current Working Groups:**
- ADAC AI & Analytics
  - Data Marketplace
  - Data Sharing
  - Master Data Management
  - Federated Data Governance
  - Data Mesh

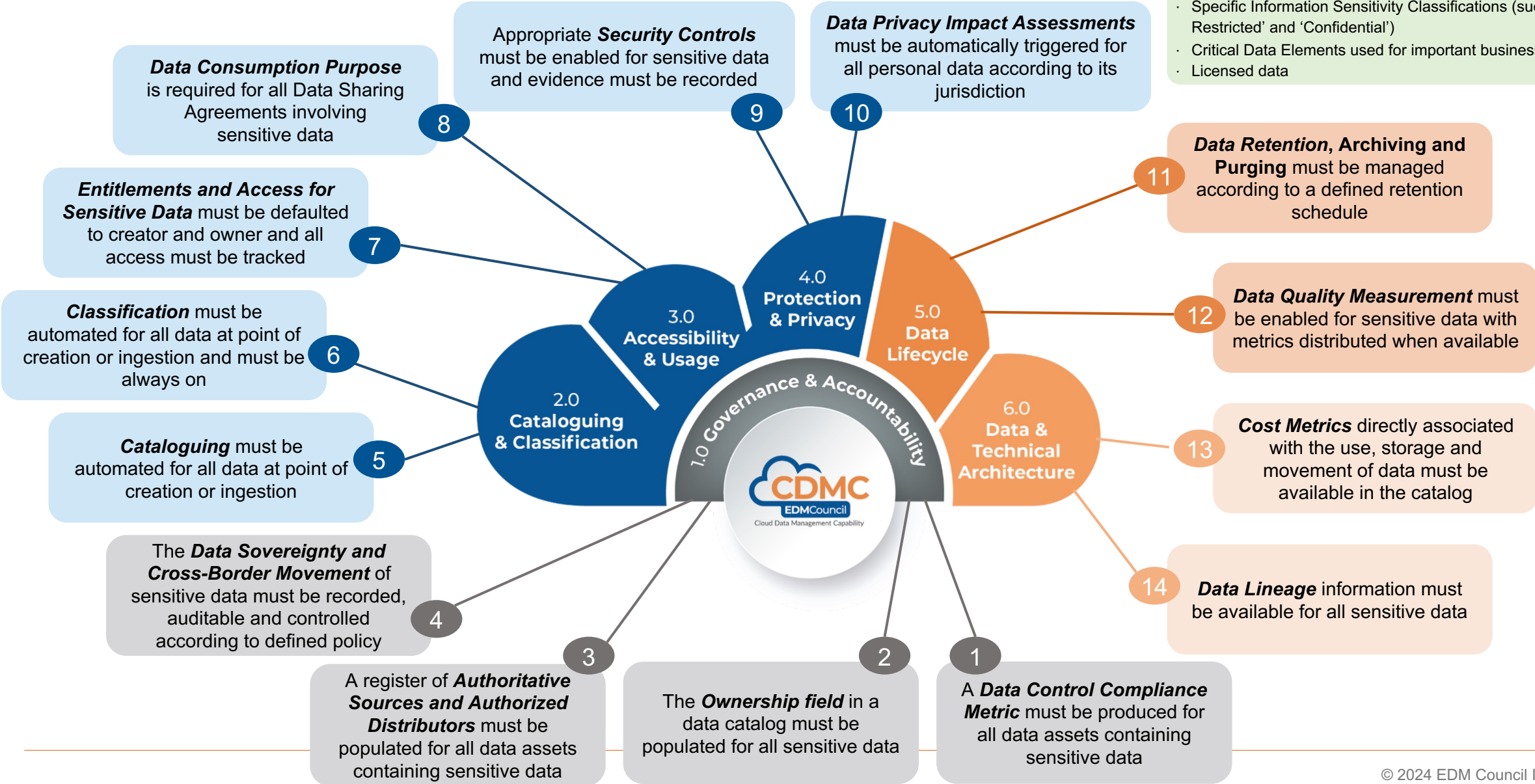
**CDMC Framework v1 Published September 2021**

# Cloud Data Management Capabilities (CDMC)

## 14 Key Controls and Automations

**Sensitive Data** includes classifications such as:

- Personal Information (PI) / Sensitive Personal Data
- Personally Identifiable Information (PII)
- Client Identifiable Information
- Material Non-Public Information (MNPI)
- Specific Information Sensitivity Classifications (such as 'Highly Restricted' and 'Confidential')
- Critical Data Elements used for important business processes
- Licensed data



# CDMC Components, Capabilities & Sub-Capabilities

Component	Capability	Sub-Capability
<b>1. Governance &amp; Accountability</b>    	<b>1.1 The Cloud Data Management business case</b> is defined and measurable	1.1.1 Cloud data management business cases are defined 1.1.2 Cloud data management business cases are governed and syndicated
	<b>1.2 Data ownership</b> established for both migrated & cloud-generated data	1.2.1 Data Owner roles and responsibilities are defined and agreed 1.2.2 Data ownership is established in the Cloud
	<b>1.3 Data sourcing and consumption</b> are governed and supported by automation	1.3.1 Data sourcing is managed and authorized 1.3.2 Data consumption is governed and supported by automation
	<b>1.4 Data Sovereignty and Cross-Border Data Movement</b> are actively managed	1.4.1 Sovereignty of data is tracked 1.4.2 Data Sovereignty and Cross-Border Data movement risks are mitigated
<b>2. Data Cataloguing &amp; Classification</b>  	<b>2.1 Data catalogues</b> are implemented, used and interoperable	2.1.1 Data cataloguing is defined, scoped and actively used 2.1.2 Metadata is discoverable, enriched, managed and exposed in Data Catalogues 2.1.3 Data catalogues are interoperable across multi and hybrid cloud environments
	<b>2.2 Data classifications</b> are defined and used	2.2.1 Data classifications are defined and approved 2.2.2 Data classifications are applied and actively used
<b>3. Accessibility &amp; Usage</b>  	<b>3.1 Data entitlements</b> are managed, enforced and tracked	3.1.1 Data entitlement rights and obligations are captured as metadata 3.1.2 Data entitlement rights are enforced 3.1.3 Automated access and entitlement tracking is established
	<b>3.2 Ethical access, use, and outcomes of data</b> are managed	3.2.1 Data Ethics organization structures are established 3.2.2 Data Ethics processes are operational
<b>4. Protection &amp; Privacy</b>  	<b>4.1 Data is secured,</b> and controls are evidenced	4.1.1 Encryption policies are defined and enforced for data at rest, in motion, and in use 4.1.2 Implementation of data security controls is evidenced 4.1.3 Data obfuscation techniques are defined, scoped and applied 4.1.4 A Data Loss Prevention program is in place
	<b>4.2 A data privacy</b> framework is defined and operational	4.2.1 A data privacy framework is defined and agreed 4.2.2 The data privacy framework is operational
<b>5. Data Lifecycle</b>  	<b>5.1 The data lifecycle</b> is planned and managed	5.1.1 A data lifecycle management framework is defined and adopted 5.1.2 The data lifecycle is implemented and managed
	<b>5.2 Data quality</b> is managed	5.2.1 Data Quality rules management is established 5.2.2 Data Quality measurement is established and operational 5.2.3 Data Quality metrics reporting is established and operational 5.2.4 Data Quality issue management is established and operational
<b>6. Data &amp; Technical Architecture</b>  	<b>6.1 Technical design principles</b> are established and applied	6.1.1 Optimization of cloud use and cost efficiency is facilitated 6.1.2 Principles for data availability and resilience are established and applied 6.1.3 Backups and point-in-time recovery are supported 6.1.4 Portability and exit planning are supported
	<b>6.2 Data provenance and lineage</b> are understood	6.2.1 Multi-environment lineage discovery is automated 6.2.2 Data lineage changes are tracked and managed 6.2.3 Data lineage reporting and visualization are implemented

# Poll #1

Who (what role) is primarily managing risks within your AI and Analytics platforms?

- **Security**
- **Privacy**
- **Technology**
- **Compliance**
- **Risk**
- **Governance**
- **Other**

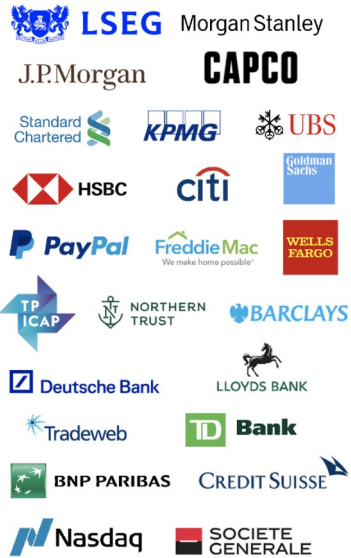


# AI, Data & Analytics Capabilities built on CDMC automated controls

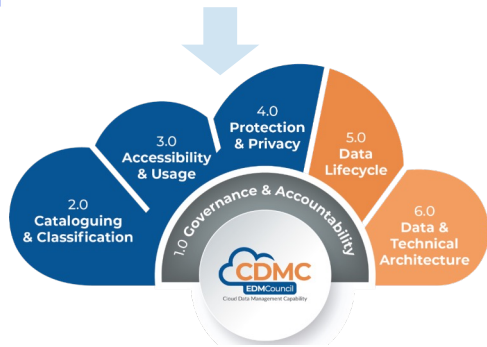
2020-2021



## CDMC Working Group



## Cloud & Technology Provider Certification



CDMC v1 Published 2021

AWS, Azure, GCP, Snowflake Certified 2021-2023

2022-2023



## Analytics Working Group

### Q4'22 - Q4'23: Extend CDMC for AI, Data & Analytics

**Members:** LSEG, Databricks, Standard Chartered, UBS, JPMorgan, Standard Bank, Schrodgers, Vanguard, Cloudera, Acquis Cortico-X

- ✓ Add **AI models & Analytical assets** to CDMC
- ✓ Add **NIST AI Risk Management Framework**
- ✓ Add **Data Ethics in Practice Framework**

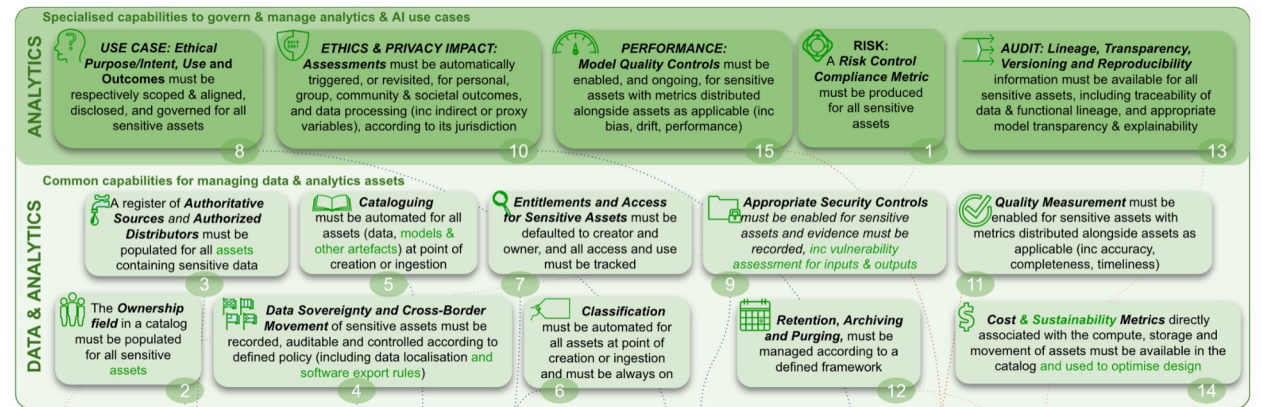
2024



## AI, Data & Analytics Controls (ADAC)

- ADAC framework - broad SME review
- ADAC Information Model & metadata inter-op
- ADAC Data Product extensions
- ADAC Regulatory mapping additions
- Reference Architecture

## ADAC Use Case Governance



## ADAC AI, Data & Analytics Capabilities

# AI, Data & Analytics Capabilities built on CDMC automated controls

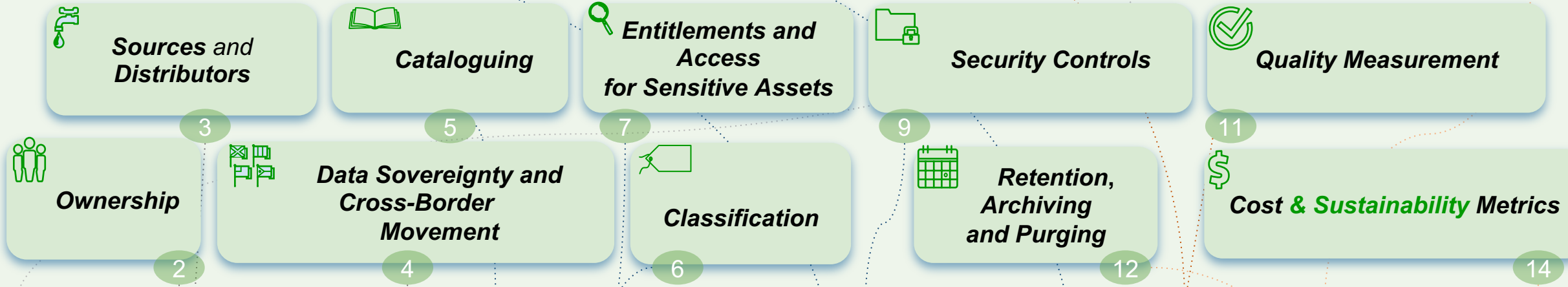
Specialised capabilities to govern & manage analytics & AI use cases

ANALYTICS

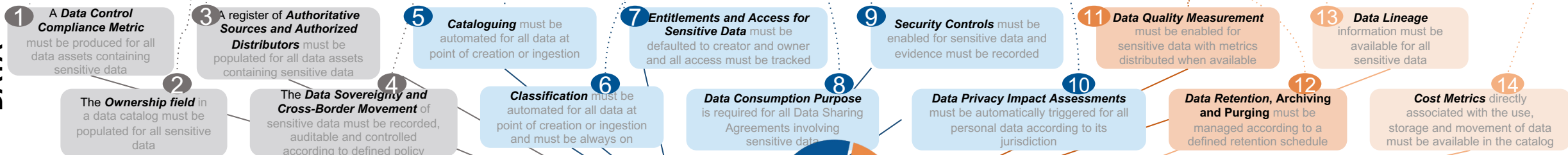


Common capabilities for managing data & analytics assets

DATA & ANALYTICS



DATA



- Classification of Sensitive Assets** includes:
- Personal Information (PI) / Sensitive Personal Data
  - Personally Identifiable Information (PII)
  - Client Identifiable Information
  - Material Non-Public Information (MNPI)
  - Specific Information Sensitivity Classifications (such as 'Highly Restricted' and 'Confidential')
  - Critical Data Elements used for important business processes
  - Licensed data
  - Model IP [new]
  - EU AI Act High Risk intended use models [new]

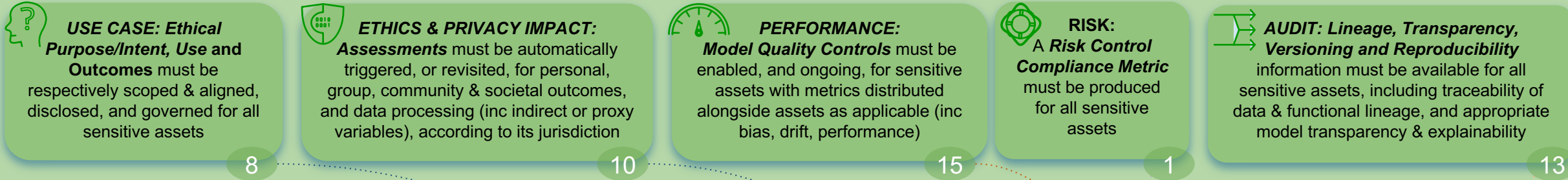


For more detail see the CDMC+ Analytics working document: [https://docs.google.com/document/d/1Cg\\_8a0RStI550epwDGk\\_IBwG6ui8fGrB/](https://docs.google.com/document/d/1Cg_8a0RStI550epwDGk_IBwG6ui8fGrB/)

# AI, Data & Analytics Capabilities built on CDMC automated controls

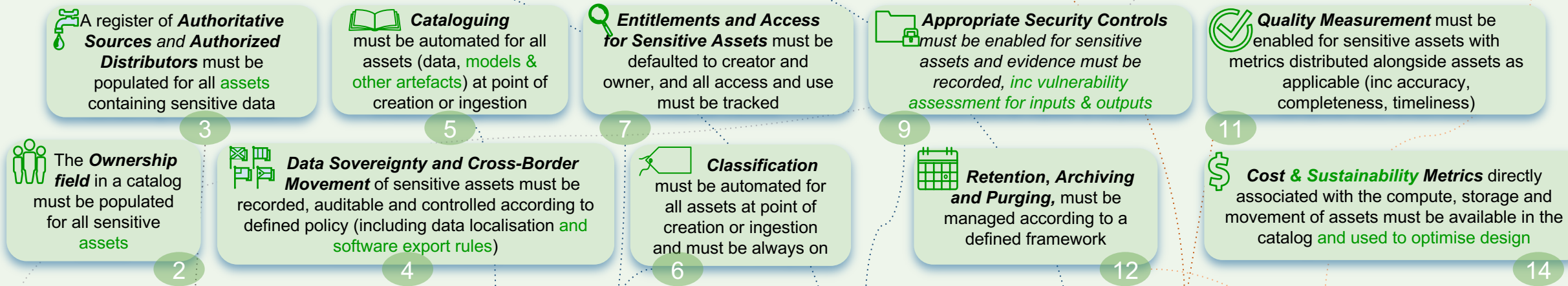
## Specialised capabilities to govern & manage analytics & AI use cases

ANALYTICS

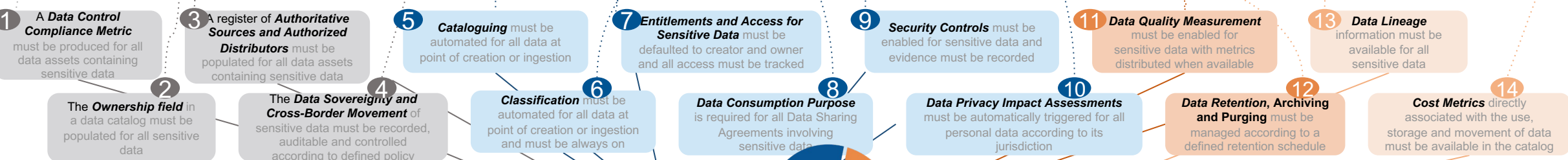


## Common capabilities for managing data & analytics assets

DATA & ANALYTICS



DATA



- Classification of Sensitive Assets** includes:
- Personal Information (PI) / Sensitive Personal Data
  - Personally Identifiable Information (PII)
  - Client Identifiable Information
  - Material Non-Public Information (MNPI)
  - Specific Information Sensitivity Classifications (such as 'Highly Restricted' and 'Confidential')
  - Critical Data Elements used for important business processes
  - Licensed data
  - Model IP [new]
  - EU AI Act High Risk intended use models [new]



For more detail see the CDMC+ Analytics working document: [https://docs.google.com/document/d/1Cg\\_8a0RStI550epwDGk\\_IBwG6ui8fGrB/](https://docs.google.com/document/d/1Cg_8a0RStI550epwDGk_IBwG6ui8fGrB/)

# CDMC AI, Data & Analytics working group - 2024 plans

## New Activity

### ADAC framework reviews

- Industry stakeholder reviews
- Is this covering the right areas and heading in the right direction?
- Does your organisation want to participate in a deep dive on writing the best practice in one or more of the 15 focus areas?
- Any other comments or opportunities?

### ADAC Spec

- For each of the 15 areas: a single page spec for the control, and automation and information model definitions.
- Built on CDMC v1.1.1

## Continuing Activity

### Regulatory Mapping

- Complete the “How does ADAC help NIST AI RMF?” reg mapping explainer sheet
- Show other reg mappings (eg EU AI act)

### Technical implementations

- Complete mapping of ADAC to popular analytics platform controls

### Data Products

- Analytics extensions for the DPRoD standard and CDMC+ Data Product spec - continue to work with those sibling groups

## GET INVOLVED

- Join our ADAC weekly working group meetings
- Mondays 1500-1600GMT
- Sign up link:

<https://forms.gle/xpkMaa4ondxNdHxb9>

## Future Plans

- Stakeholder interest map
- Reference Architecture
- Analytics storytelling (Q2)
- +Additional topics from new members.

# AI, Data & Analytics Capabilities built on CDMC automated controls

## Specialised capabilities to govern & manage analytics & AI use cases

ANALYTICS



**USE CASE: Ethical Purpose/Intent, Use and Outcomes** must be respectively scoped & aligned, disclosed, and governed for all sensitive assets

8



**ETHICS & PRIVACY IMPACT: Assessments** must be automatically triggered, or revisited, for personal, group, community & societal outcomes, and data processing (inc indirect or proxy variables), according to its jurisdiction

10



**PERFORMANCE: Model Quality Controls** must be enabled, and ongoing, for sensitive assets with metrics distributed alongside assets as applicable (inc bias, drift, performance)

15



**RISK: A Risk Control Compliance Metric** must be produced for all sensitive assets

1



**AUDIT: Lineage, Transparency, Versioning and Reproducibility** information must be available for all sensitive assets, including traceability of data & functional lineage, and appropriate model transparency & explainability

13

## Common capabilities for managing data & analytics assets

DATA & ANALYTICS



A register of **Authoritative Sources and Authorized Distributors** must be populated for all **assets** containing sensitive data

3



**Cataloguing** must be automated for all assets (data, **models & other artefacts**) at point of creation or ingestion

5



**Entitlements and Access for Sensitive Assets** must be defaulted to creator and owner, and all access and use must be tracked

7



**Appropriate Security Controls** must be enabled for sensitive assets and evidence must be recorded, **inc vulnerability assessment for inputs & outputs**

9



**Quality Measurement** must be enabled for sensitive assets with metrics distributed alongside assets as applicable (inc accuracy, completeness, timeliness)

11



The **Ownership field** in a catalog must be populated for all sensitive **assets**

2



**Data Sovereignty and Cross-Border Movement** of sensitive assets must be recorded, auditable and controlled according to defined policy (including data localisation and **software export rules**)

4



**Classification** must be automated for all assets at point of creation or ingestion and must be always on

6



**Retention, Archiving and Purging**, must be managed according to a defined framework

12



**Cost & Sustainability Metrics** directly associated with the compute, storage and movement of assets must be available in the catalog and **used to optimise design**

14

[CLICK HERE](#)  
FOR MORE  
INFORMATION & TO  
JOIN OUR ADAC SIG:



# Questions?

**EDM** Webinar 



# LSEG

Operating in every  
major financial center

170 Countries

45,000 Customers

500 Exchange Feeds

150,000 Data Sources

150 Billion market events per day

17 Million messages per second

Powering the  
financial markets

Data & Feeds

Workflows & News

Financial Analytics

Benchmarks & Indices

Equity & FX Venues

Clearing & Reporting

Screening & Identity



**LSEG** DATA &  
ANALYTICS



**FTSE  
RUSSELL**  
An LSEG Business



**LONDON  
STOCK  
EXCHANGE**  
An LSEG Business



**LSEG FX**



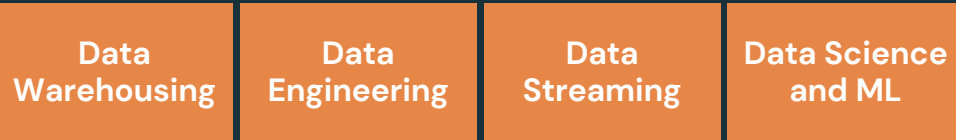
**LSEG** POST  
TRADE



**LSEG** RISK  
INTELLIGENCE



## databricks Data Intelligence Platform



**Unity Catalog**  
Fine-grained governance for data and AI

**Delta Lake**  
Data reliability and performance

**Cloud Data Lake**  
All structured and unstructured data



# Databricks Data Intelligence Platform

## Simple

Unify your data warehousing and AI use cases on a single platform.

## Open

Built on open source and open standards.

## Secure

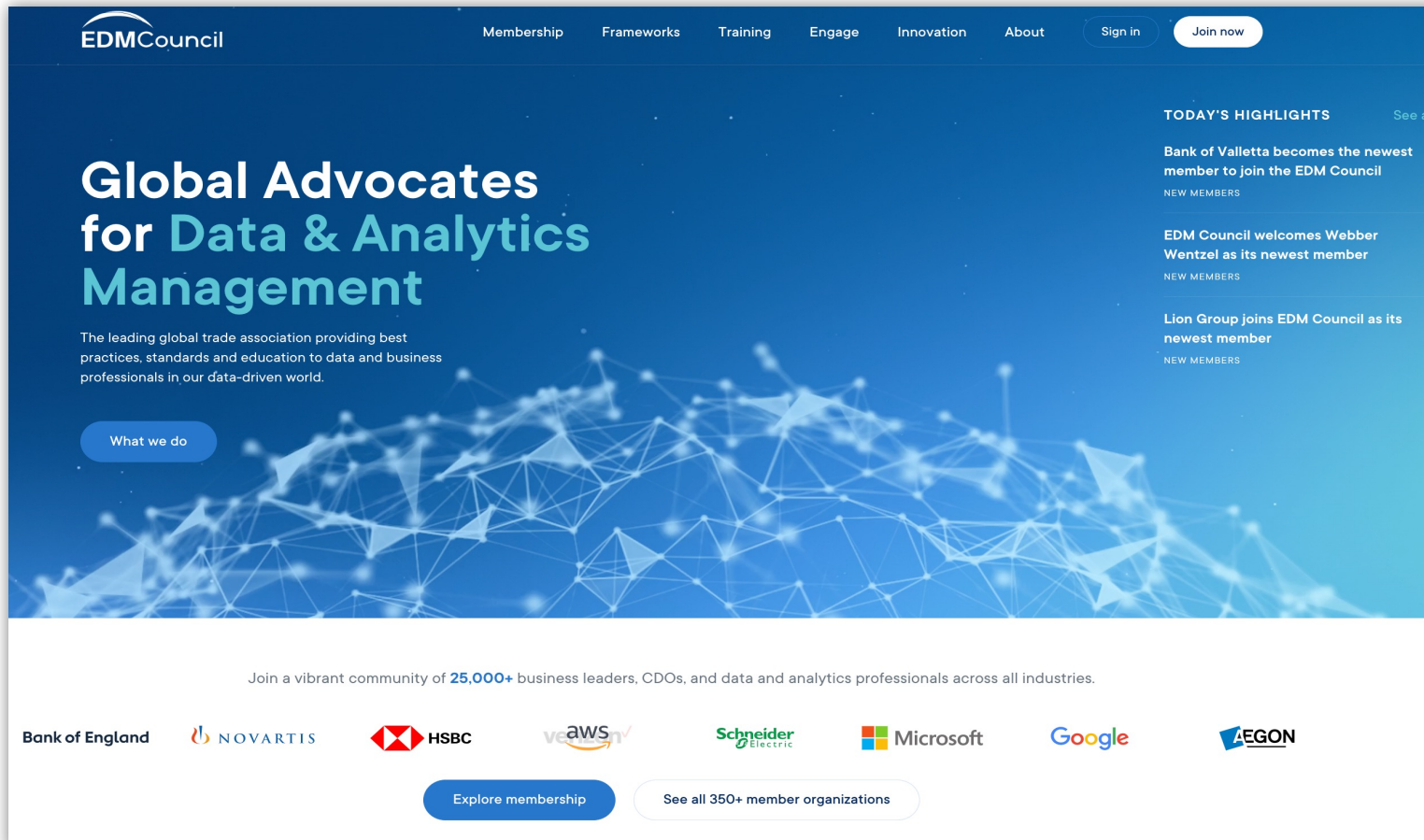
Unified governance and fine-grain controls.

## Multi-cloud

One consistent data platform across clouds.



# 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](https://edmcouncil.org)



# EDM Webinar

## Thank you!

[CLICK HERE](#) FOR MORE INFORMATION  
& TO JOIN OUR ADAC SIG:



  
**EDM Council**

