Reimagining Data Governance in the Cloud

A conversation with

Oli Bage
Head of Architecture, Data & Analytics
LSEG

Mojgan Ahmadi
Principal Solutions Architect
AWS

Jen Code
Principal Cloud Governance
AWS

Christopher Glover
Field CTO
BigID
Moderated by **Mike Meriton**
Co-Founder & COO, EDM Council

- Joined EDM Council full-time 2015 to lead Industry Engagement
- EDM Council Co-Founder & First Chairman (2005-2007)
- Former CEO GoldenSource (2002-2015)
- Former Executive for D&B Software and Oracle
- FinTech Innovation Lab – Executive Mentor (2011 – Present)
Customer Need
Cloud Data Management Capability (CDMC)

CDMC Working Group

Cloud Challenges

• **Inefficiency**: data, technology, regulatory and planning challenges on nearly every cloud implementation
• **93% of firms** use 2 or more cloud providers*

CDMC Group Objectives

1. Define consistent best practices for a hybrid-cloud world
2. Align key cloud data controls to meet regulatory obligations for Sensitive Data
3. Accelerate Cloud Adoption with comprehensive framework

* Source: Flexera, 2020 *State of the Cloud*
CDMC: Industry Engagement

70+ Leading firms and 226 participants actively participating since May 2020

CDMC Working Group

- LSEG
- Morgan Stanley
- J.P. Morgan
- CAPCO
- Standard Chartered
- KPMG
- UBS
- HSBC
- Citi
- Goldman Sachs
- PayPal
- Freddie Mac
- Wells Fargo
- TP ICAP
- Northern Trust
- Barclays
- Deutsche Bank
- Lloyds Bank
- Tradeweb
- Traiana
- BNP Paribas
- Credit Suisse
- Nasdaq
- Société Générale

Cloud & Technology Provider Certification

- AWS
- Microsoft Azure
- Google Cloud
- IBM Cloud
- Snowflake
- Collibra
- Informatica
- BigID
- data.world
- Privitar

Regulatory Engagement

- US: Federal Reserve, OCC, FDIC, NCUA, NAIC
- Canada: OSFI, BoC, CDIC
- UK: BoE, FCA, IOC
- EU: EBA, ECB, DORA Act
- Germany: BaFin
- Swiss: FinMA
- Japan: FSA
- Australia: APRA
- Singapore: MAS
- India: RBI, SEBI
- Africa/Middle East regulators
- Others

Go-to-market Support

- Training Courses
- Cloud Service Certification
- Open Source Tools
- CDMC Authorized Partner Program

2H 2021 – 1H 2022

Other Industries

- Life Sciences
- Telecommunications
- Manufacturing
- Retail / Services
- Consumer Tech
- Government/Defense
- Others

CDMC Working Group Publication on track for Q3 2021
## CDMC Capabilities: Requirements & Automated Controls

<table>
<thead>
<tr>
<th>Component</th>
<th>Capability</th>
<th>Sub-Capability</th>
<th>CDMC Automation Checklist (DRAFT)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1. Governance &amp; Accountability</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.1 The Cloud Data Management business case is defined and measurable</td>
<td>1.1.1 Cloud data management business cases are governed and syndicated</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.1.2 Cloud data management business cases are defined and include a value realization framework</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.2 Data ownership established for both migrated &amp; cloud-generated data</td>
<td>1.2.1 Data Owner roles and responsibilities are defined and agreed</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.2.2 Data ownership is established in the Cloud</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.3 Data sourcing and consumption are governed and supported by automation</td>
<td>1.3.1 Data sourcing is managed and authorized</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.3.2 Data consumption is governed and supported by automation</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.4 Data Sovereignty and Cross-Border Data Movement are actively managed</td>
<td>1.4.1 Jurisdictional location of data is tracked and reported</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.4.2 Data Sovereignty and Cross-Border Data movement risks are mitigated</td>
<td></td>
</tr>
<tr>
<td><strong>2. Data Cataloguing &amp; Classification</strong></td>
<td>2.1 Data catalogues are implemented, used and interoperable</td>
<td>2.1.1 Data cataloguing is defined, scoped and actively used</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.1.2 Metadata is discoverable, enriched, managed and exposed in Data Catalogues</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.1.3 Data catalogues are interoperable across multi and hybrid cloud environments</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.2 Data classifications are defined and used</td>
<td>2.2.1 Data classifications are defined and approved</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.2.2 Data classifications are applied and actively used</td>
<td></td>
</tr>
<tr>
<td><strong>3. Accessibility &amp; Usage</strong></td>
<td>3.1 Data entitlements are managed, enforced and tracked</td>
<td>3.1.1 Data entitlement rights and obligations are captured as metadata</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.1.2 Data entitlement rights are enforced</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.1.3 Enforcement of data entitlement rights is evidenced</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.2 Ethical access, use, and outcomes of data are managed</td>
<td>3.2.1 Data Ethics organization structures are established</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.2.2 Data Ethics processes are operational</td>
<td></td>
</tr>
<tr>
<td><strong>4. Protection &amp; Privacy</strong></td>
<td>4.1 Data is secured, and controls are evidenced</td>
<td>4.1.1 Encryption policies are defined and enforced for data at rest, in motion, and in use</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.1.2 Implementation of data security controls is evidenced</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.1.3 Data obfuscation techniques are defined, scoped and applied</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.1.4 A Data loss Prevention regime is in place</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4.2 A data privacy framework is defined and operational</td>
<td>4.2.1 A data privacy framework is defined and agreed</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4.2.2 The data privacy framework is operational</td>
<td></td>
</tr>
<tr>
<td><strong>5. Data Lifecycle</strong></td>
<td>5.1 Data quality is managed</td>
<td>5.1.1 Data Quality rules management is established</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.1.2 Data Quality measurement is established and operational</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.1.3 Data Quality metrics reporting is established and operational</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.1.4 Data Quality issue management is established and operational</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5.2 The data lifecycle is planned and managed</td>
<td>5.2.1 Data archiving and purging are managed</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5.2.2 TBA</td>
<td></td>
</tr>
<tr>
<td><strong>6. Data &amp; Technical Architecture</strong></td>
<td>6.1 Data provenance and lineage are understood</td>
<td>6.1.1 Multi-environment lineage discovery is automated</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.1.2 Data provenance tracking and change management</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.1.3 Data lineage reporting and visualization</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.2.1 Backups and point-in-time recovery are supported</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.2.2 Optimization of cloud use and cost efficiency is facilitated</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.2.3 Portability &amp; exit planning are supported</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.2.4 Principles for data availability and resiliency are established and applied</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6.2 Technical design principles are established and applied</td>
<td>6.2.1 Data provenance and lineage are understood</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.2.2 Optimization of cloud use and cost efficiency is facilitated</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.2.3 Portability &amp; exit planning are supported</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>6.2.4 Principles for data availability and resiliency are established and applied</td>
<td></td>
</tr>
</tbody>
</table>
Cloud Data Management Capabilities (CDMC)
14 Key Controls for Managing Data Risk

1. A Data Control Compliance Metric derived from all CDMC Key Controls must be produced for all data assets containing sensitive data

2. The Ownership field in a data catalog must be populated for all sensitive data

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

4. The Data Sovereignty and Cross-Border Movement of sensitive data must be recorded, auditable and controlled according to defined policy

5. Cataloguing must be automated for all data at point of creation or ingestion

6. Classification must be automated for all data at point of creation or ingestion and must be always on

7. Entitlements and Access for Sensitive Data must be defaulted to creator and owner and access must be tracked for all sensitive data

8. Data Consumption Purpose must be provided for all Data Sharing Agreements involving sensitive data

9. Appropriate Security Controls must be enabled for sensitive data and evidence must be recorded

10. Data Privacy Impact Assessments must be automatically triggered for all personal data according to its jurisdiction

11. Data Quality Measurement must be enabled for sensitive data with metrics distributed when available

12. Data Retention, Archiving and Purging must be managed according to a defined retention schedule

13. Data Lineage information must be available for all sensitive data

14. Cost Metrics directly associated with the use, storage and movement of data must be available in the catalog

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 Regulatory Engagement Update

• Objective
  • Engage Global Regulators for CDMC awareness, engagement and feedback

• Key Regulatory Updates
  • Formed CDMC Regulatory Engagement Committee
  • Significant Momentum in Engaging Regulators

• Completed 3 Initial CDMC Regulatory Briefings
  • US Federal Reserve Board
  • Bank of England
  • APRA (Australia)
  • OSFI (Canada)
  • ICO (Information Commissions Office UK)

• Upcoming Initial CDMC Regulatory Briefings
  • African Regulators (20) Cloud Symposium (July)
  • 2nd annual meeting hosted by Standard Bank

• Scheduling CDMC Deeper Dive Sessions
  • Bank of England
  • Federal Reserve Board (potential OCC addition)
  • APRA

• Additional Regulatory Outreach support from EDMC CDMC Partners
  • IBM/Promontory, PwC, UBS, Standard Bank, others…
CDMC Deliverables Availability

- **July 2021**: 14 CDMC Key Controls & Automations
  - First CDMC Industry Deliverable published 7 July 2021
  - Auditable, evidence-based Key CDMC Controls & Automations for managing and protecting sensitive data
  - 20-page document
  - Summarizes and elaborates on key controls in Framework

  **DOWNLOAD 14 CMDC KEY CONTROLS:**
  [https://edmcouncil.org/page/cdmc-14-key-controls-and-automation](https://edmcouncil.org/page/cdmc-14-key-controls-and-automation)

- **September 2021**: Complete CDMC Framework
  - Comprehensive 150+ page document
  - 6 Components, 14 capabilities, 37 sub-capabilities
  - Incorporates the 14 Key Controls & Automations
  - Availability of 2-day training course
  - Other Online Digital Support tools
## CDMC: Accelerating Trusted Cloud Adoption

<table>
<thead>
<tr>
<th>CDMC Industry Objective:</th>
<th>Companies</th>
<th>Cloud Service Providers (CSP)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Build Trust, Confidence, and Dependability for Cloud Adoption</strong></td>
<td>CDMC structured framework of auditable Cloud processes and controls – especially for sensitive data</td>
<td>CDMC requirements and controls can be automated into CSP platforms which accelerates adoption and provides market confidence</td>
</tr>
</tbody>
</table>

### Application, Technology & Data Providers
Incorporates certified CDMC capabilities and controls into services and solutions to ensure high degree of reliability and operational effectiveness

### Consultants & Systems Integrators (SI)
CDMC enables training & assessments, gap analysis, strategy development, and execution services for end clients adopting cloud

### Regulators
CDMC provides industry guidance for auditing and validating key cloud controls, especially for sensitive data
Cloud Service Provider Perspective
Cloud Data Management Capabilities (CDMC) 14 Key Controls for Managing Data Risk

Control Enabled by:

Company + CSP + Tech Provider
Cloud Service Provider
Technology Provider

1. A Data Control Compliance Metric derived from all CDMC Key Controls must be produced for all data assets containing sensitive data

2. The Ownership field in a data catalog must be populated for all sensitive data

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

4. The Data Sovereignty and Cross-Border Movement of sensitive data must be recorded, auditable and controlled according to defined policy

5. Cataloguing must be automated for all data at point of creation or ingestion

6. Classification must be automated for all data at point of creation or ingestion and must be always on

7. Entitlements and Access for Sensitive Data must be defaulted to creator and owner and access must be tracked for all sensitive data

8. Data Consumption Purpose must be provided for all Data Sharing Agreements involving sensitive data

9. Appropriate Security Controls must be enabled for sensitive data and evidence must be recorded

10. Data Privacy Impact Assessments must be automatically triggered for all personal data according to its jurisdiction

11. Data Quality Measurement must be enabled for sensitive data with metrics distributed when available

12. Data Retention, Archiving and Purging must be managed according to a defined retention schedule

13. Data Lineage information must be available for all sensitive data

14. Cost Metrics directly associated with the use, storage and movement of data must be available in the catalog

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
Technology Vendor Perspective
Cloud Data Management Capabilities (CDMC)  
14 Key Controls for Managing Data Risk

Control Enabled by:
- Company + CSP + Tech Provider
- Technology Provider

1. A Data Control Compliance Metric derived from all CDMC Key Controls must be produced for all data assets containing sensitive data

2. The Ownership field in a data catalog must be populated for all sensitive data

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

4. The Data Sovereignty and Cross-Border Movement of sensitive data must be recorded, auditable and controlled according to defined policy

5. Cataloguing must be automated for all data at point of creation or ingestion

6. Classification must be automated for all data at point of creation or ingestion and must be always on

7. Entitlements and Access for Sensitive Data must be defaulted to creator and owner and access must be tracked for all sensitive data

8. Data Consumption Purpose must be provided for all Data Sharing Agreements involving sensitive data

9. Appropriate Security Controls must be enabled for sensitive data and evidence must be recorded

10. Data Privacy Impact Assessments must be automatically triggered for all personal data according to its jurisdiction

11. Data Quality Measurement must be enabled for sensitive data with metrics distributed when available

12. Data Retention, Archiving and Purging must be managed according to a defined retention schedule

13. Data Lineage information must be available for all sensitive data

14. Cost Metrics directly associated with the use, storage and movement of data must be available in the catalog

EDMCouncil
Creating Consistent Data Capabilities across Environments

- Design for Privacy, Security & Governance
- Must span all clouds
- Enforce consistency
- Minimization principle
- Resilience
- Diverse controls for different actions
Questions?

BigID

EDM Webinar
Data Management Reimagined

Extensible, Open Platform for Data Visibility & Control in Privacy, Security & Governance

Action Your Data

Know Your Data

Connect Your Data

Catalog • Classify • Correlate • Cluster

BigID App Marketplace for privacy, security, and governance controls

Discovery-in-Depth for deeper data visibility, insight, and inventory

Files • Images • Unstructured • Structured • Big Data • NoSQL • Pipeline • SaaS • IaaS • Apps • Messaging • Mainframe
FOR MORE INFORMATION:

BigID
https://bigid.com/contact/
info@bigid.com