Data Management and Governance

What are they?

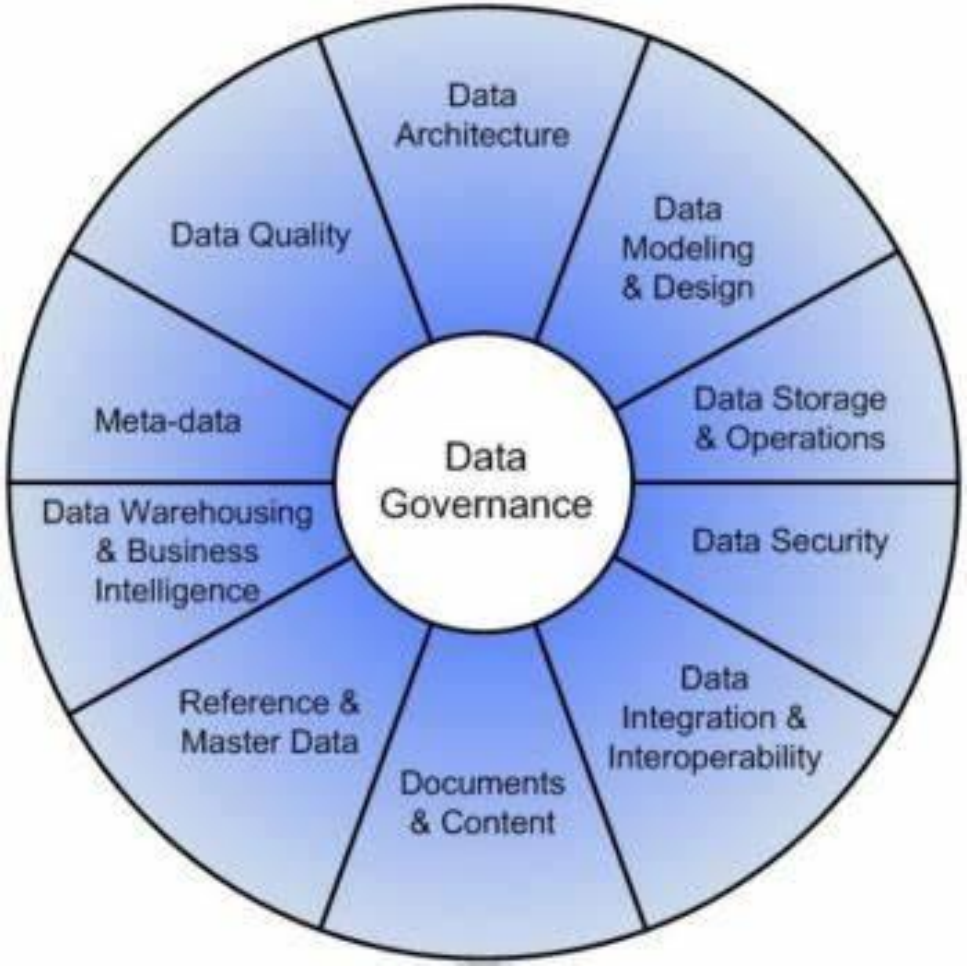
What's the difference?

Case study

Dr Karolyn Kerr
CDAO
August 2022

Data Management – What is it?

- Data management from the DM Body of Knowledge
- Data Governance is at the centre – you can't have one without the other
- The scope encompasses some of the work of IT, but also business intelligence



Governance

Governance is key. The entire organisation needs to embrace change to the way they work, where data is valued rather than an output of a transaction. The data management function cannot deliver without governance - decisions need to be made by the leaders of the business units.

Layers of governance are needed. Data stewards are given the role of ensuring the principles agreed by the data governance group and decisions made are implemented. These are people that work in the business units and know the data and its context. They will make the operational decisions based on the strategic directives given by the data governance group. If they have any issues, these are raised to the data governance group for a decision.

Not all the work programme can be addressed at once - prioritisation is required. This will consider the data most used for critical decision making - your master data.

What's the difference between Data Management and Data Governance?

Maintaining the Data Architecture

Data governance

- The executive approve the target conceptual data architecture, ensuring it aligns with strategic plans

Data management

- Develops the conceptual data architecture and associated logical and physical data models
- Maintains the integrity of the data models over time to ensure changes align with the agreed target conceptual architecture
- Uses data management tools to understand the impact of changes to data models will have on reporting
- Enables access to data models to stakeholders so they know what data is available and its relationships

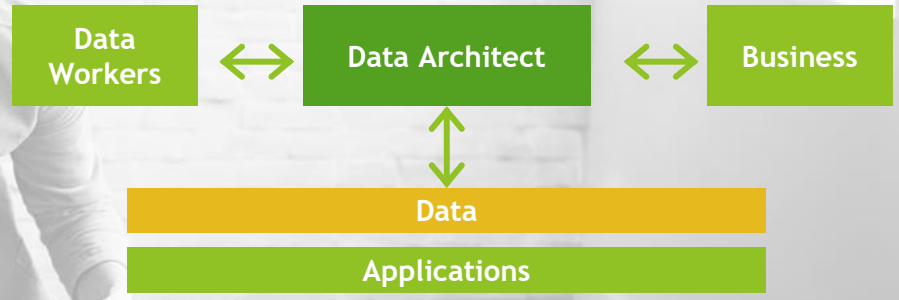
Conceptual data architecture is rarely done. The benefit is that it provides a clear picture of what data means to the organisation for the executive to the data management team on decisions around data structures to ensure we build for the future, not just the present.

The data architect is one of the most important roles for operationalizing the data strategy

The data architect:

- Acts as a **"translator"** between the business and data workers to communicate data and technology requirements.
- **Facilitates** the creation of the data strategy.
- **Manages** the enterprise data model.
- Has a greater **knowledge** of operational and analytical data use cases.
- **Recommends** data management policies and standards and maintains data management artifacts.
- **Reviews** project solution architectures and identifies cross impacts across the data lifecycle.
- Is a **hands-on expert** in data management and warehousing technologies.
- Is not necessarily its own designated position but a role that can be completed by a variety of Data professionals.

Data architects bridge the gap between strategic and technical requirements.



The **data architect** must maintain a comprehensive view of the organization's rapidly proliferating data.

What's the difference between Data Management and Data Governance?

Enabling Access

Data governance
- enables access to data by making clear decisions around who is allowed access

Data Management
- Enacts the data governance decisions through metadata classifications – automating access to those with the privileges
- Looks for personally identifiable information held across the business and removes it where it's not necessary or held in insecure places, enabling the opening up of files

IT
- Reflect the agreed access protocols in systems and access profiles

Many organisations close off access to all information by default because they don't know what's where and who should be allowed to see it. A lot of institutional knowledge isn't used because no one can access it.

What's the difference between Data Management and Data Governance?

Organisation wide participation

Data governance

- requires all of the organisation to participate in the decisions around data, its management and use. This is done through layers of governance so the executive team make strategic decisions and data stewards implement them. Data stewards make operational decisions. Accountability and responsibilities are clear

Data management

- Supports the executive team by providing expertise, raising appropriate and strategic decisions to them
- Supports the data stewards to do their job – data management tools and processes that enable easy management of metadata, automated as much as possible

IT

Many organisations believe IT is responsible and accountable for making decisions around data. In reality, only the business knows the context of the data; crucial to data quality.

What's the difference between Data Management and Data Governance?

Organisation wide consistency

Data governance

- Makes decisions on contentious issues and disagreement across business units around consistency

Data management

- Supports the governance group by providing expertise, raising issues unresolved by data stewards
- Supports the data stewards to do their job – consistency through agreed commonly used business terms in the data glossary, developing a data dictionary, classifications and coding, data quality reports

The single most effective initiative is the **business glossary**. This requires not only ensuring everyone agrees to standard definition for a business term, but also that it is used in all reporting, and term definitions are accessible to users

What's the difference between Data Management and Data Governance?

Prioritisation

Data governance

- The executive team agree on the priority for large scale projects and ensures they are funded, agrees on what is the master data set

Data management

- Supports the executive team (and/or the programme management office) by providing expertise, raising decisions to them on which initiatives deliver the best ROI, are critical to the business, solves a problem that is blocking growth
- Identifies master data and reference data to enable focused improvements on the most important data

You can't manage all your data so decide what's actually used beyond counting transactions and important. Think about your core business – what are the *enduring questions* that you will always want to measure?

What's the difference between Data Management and Data Governance?

Management of Data Quality

Data governance

- The executive approve the data quality framework approach and dimensions

Data management

- Develop the data quality framework and its dimensions, characteristics and criteria with business SME's
- Develop statistical process control measures to understand if master data is within expected values
- Develop regular data quality reporting to identify persistent data quality issues, provide feedback to stewards, know what data is best suited to specific reports, and include caveats in reports

The data quality framework enables further prioritisation. Use at a more granular level for critical master data, highlights where the same dimension is an issue across all systems or datasets, provides measurements at a point in time that can be reassessed following improvement initiatives

What's the difference between Data Management and Data Governance?

Management of the Data Environment

Data governance

- The executive approve the strategic direction for data use and funding for data infrastructure

Data management

- Develop data flows from source systems into Snowflake and/or the staging layer and maintain them
- Manage the constraining of data into the data warehouse by identifying what's required for integrated reporting

The content of the data warehouse must be carefully managed to ensure only the data required for reporting is added. This reduces the resources required for data management and highlights what data is master data. This is iterative, more data will be added over time based on new reporting requirements

Case Study

Introduction

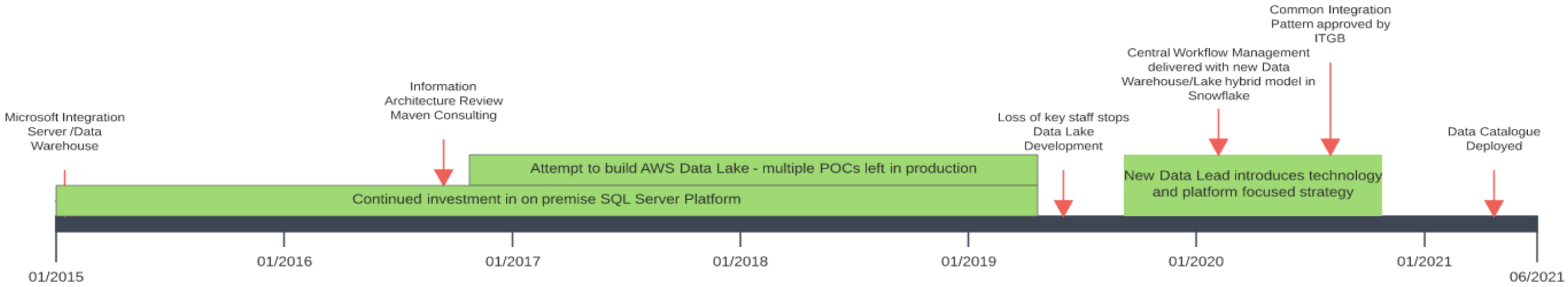
IHCs Data & Integration Platform and Data Management Practice should:

- Contribute to strategic outcomes at a reasonable cost and within realistic timeframes
- Prioritise regulatory compliance and minimise risk
- Reduce the ongoing accumulation of technical debt
- Provide a Governance Framework and controls that facilitate further maturity gains
- Manage the data pipeline from the source system through to the delivery of integrated reporting
- Enable everyone in the organisation to understand their role and accountabilities in the collection, management, and use of data

Background – Microsoft Platform

Two attempts at evolving Data Maturity at IHC have been made in the past 4-5 years with limited success.

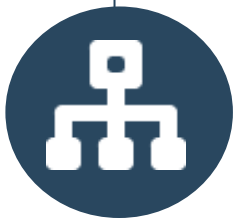
Initially a Microsoft Product Stack for reporting and Integration was built and has been in use for a number of years (approx. 2015) but has not delivered value due to being poorly governed, lacking architectural patterns or planning and a culture that rewarded delivering tactical solutions with high maintenance costs. The result has been low value data assets built directly from application datastores being used for reporting and integration. This approach has a high cost to maintain and appeared robust only at the cost of manual monitoring, multiple remediations, low trust in data quality and a huge burden on Report Developers to interpret source data with no prior modelling.



Current Environment – Prevailing Themes

Prevailing Data-Management-Related Issues:

Data Availability



Disparate data sources have made it difficult for users to gain timely access to data. As a result, users spend substantial time consolidating data and writing complex queries. There are no controls or documentation

Data Stewardship and Ownership

The ownership and stewardship of data assumed to be the responsibility of IT. Quality issues are fixed by expensive after the fact manual processes and work arounds

Data Quality Management

There are multiple data quality issues as result of not only incorrect data capture and manual integration but also misunderstandings about data terms. Key technical decisions are made at business unit level and prioritise short term time/cost savings by adding long term debt

Undefined and Complex Terminology

There are several versions and interpretations of key terms by the different data user. Significant time is spent designing process and systems to handle edge cases resulting in needless complexity

Data Strategy Objectives

The objective statement for the 2021 data strategy is:

> To develop IHCs Data Platform, People and Assets to allow data to be transformed into information that can improve the lives of people with intellectual disabilities

We will do this by:

- Ensuring sufficient controls & processes are in place to meet all regulatory requirements
- Changing the organisation wide view of data from the by product of business process to being a valuable and quantifiable asset
- Reducing the decentralized and project focused technology decision making that is producing technical debt in the data domain
- Reorganizing Data Team roles and responsibilities to ensure key people are accountable and empowered to deliver
- Maturing the organization's data literacy by promoting the use of modelled and curated data within a governed framework
- Defining ownership and stewardship of the organization’s various data assets, including delegating authority to Data Stewards
- Defining, measuring and reporting on Data Quality
- Producing a catalogue that quantifies and measures the value of all IHC’s data assets by linking them to KPIs and Metrics

Data Strategy Guiding Principles

IHCs Data Strategy guiding principles are:

Principle #1 Data is owned by the Business Owner of the system/process which creates it
Principle #2 Data is centrally integrated using data integration tools not manual input
Principle #3 Data owners and data stewards are accountable and responsible for their domains
Principle #4 Quality of data will be defined, measured, maintained, and managed
Principle #5 Data definitions are consistent, versioned and approved by business owners they are maintained in a Central Catalogue
Principle #6 Lifecycle management is implemented for data based on it's classification
Principle #7 Data is appropriately secured across its lifecycle using defined standards
Principle #8 Data governance policies and procedures once approved apply to everyone

Operational Drivers

The Data Strategy is a mechanism for delivering organisational outcomes, therefore the initial key business drivers and their order of priority are:

1. **Risk and Compliance Management (including Lifecycle Management)**
2. **Pipeline & Stakeholder Management**
3. **Operational Excellence and Efficiency**

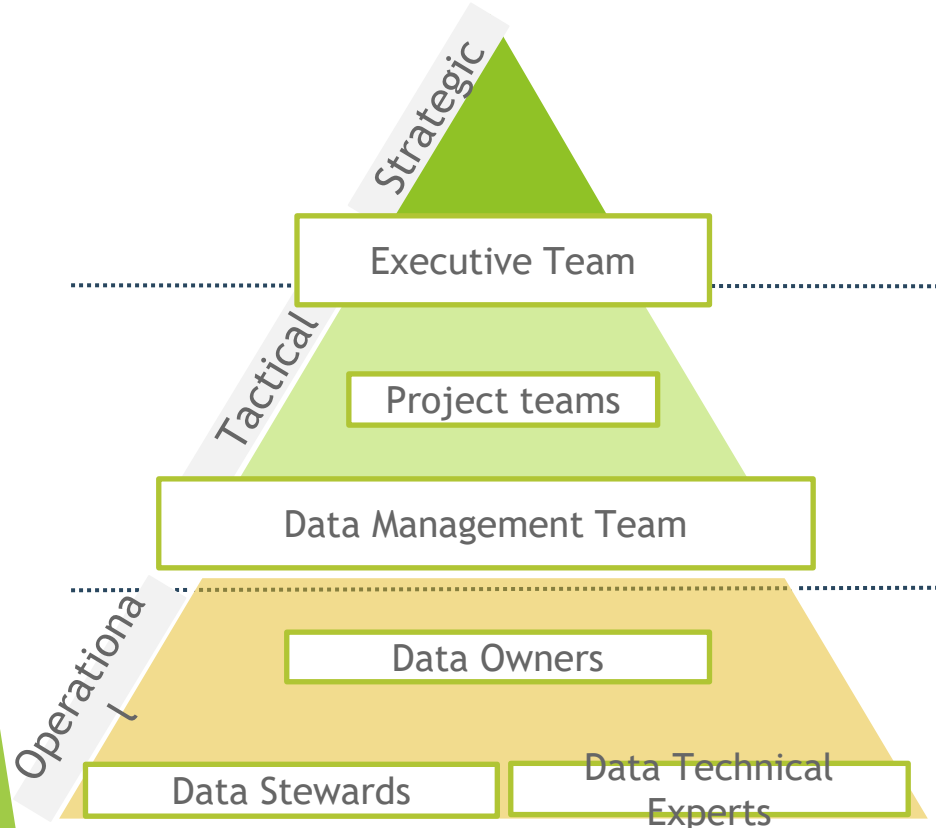
Risk & Compliance Management

1 Implement Information Life Cycle Management

There is an urgent need to prioritise data classification and retention/removal at IHC to reduce the risk or breaches and optimise the value of data assets

Roles and Organisational Structure: Key Data Roles and Committees

Proposed Data Governance Organisational Structure for IHC - An additional Governance Layer is Required



- Generate awareness and cultural change within the organization.
- Final decision authority for unresolved data conflicts.
- Approve responsibility changes and fund data improvement initiatives

- Continuously monitor progress of current data initiatives.
- Provide high-level integration between technology and the organisation.
- Projects for specific data initiatives.
- Facilitate cross-functional data initiatives.
- Support operations to do their job.

- Identify data-related pains affecting the organization.
- Recommend solutions.
- Participate in defining data rules and linking to business glossary.
- Assist in creating data policies, standards, and procedures.
- Deliver on data-related initiatives.
- Document and maintain metadata for their business unit systems.