



Expertise  
and insight  
for the future

Jukka-Pekka Paananen

# A New Data Governance Model for the Bank of Finland

Metropolia University of Applied Sciences

Master's Degree

Programme in Business Informatics

Thesis

29.11.2020

Author	Jukka-Pekka Paananen
Title	A New Data Governance Model for the Bank of Finland
Number of Pages	59 pages
Date	29 November 2020
Degree	Master's Degree
Degree Programme	Master's Degree Programme in Business Informatics
Instructors	Antti Komonen, Project Manager Dr. Thomas Rohweder, Principal Lecturer
<p>A common problem in the enterprise environment is that data is fragmented and in silos. The case company of this thesis, the Bank of Finland has a classical data governance model that suffers from similar problems. The solution proposed in this study is to modernize the data governance model.</p> <p>Both qualitative and quantitative research approach is used in this study. Formal and informal discussions were used to analyze the current state. Data collection was used to identify patterns and create the proposal. The study started with identifying the business challenge. The challenge is how to best utilize data assets and remove the usage barriers. The study was conducted by investigating the current state based on discussions with stakeholders, literature study and participation in conferences to find best practices. A proposal was formed based on the identified best practices.</p> <p>The proposal includes four pillars for data governance. These pillars are roles and responsibilities (people), processes, policies, methodology (framework) and technologies. First, Data as a Service is seen as a modern way to access data. It provides a centralized access to relevant data assets in a single location. Data factory architecture provides such a common data service model. Second, the people part includes the organizational structure and roles. Modern data governance includes data engineers, data stewards, data consumers, and data governance council members, who are the data owners. Third, for data governance process, data definition, structure and semantics need to be agreed as a common language. Data quality and integrity is a major topic to ensure reliable data usage. Change management is key to building an agile data governance process. Fourth, technology for data governance includes enterprise data management architecture, data modelling and integration, master and reference data management, and data catalogue and data portal.</p>	
Keywords	Data governance, data stewards, data engineer

# Table of Contents

1	Introduction	4
1.1	Case Company	5
1.2	Business Challenge, Objective and Outcome of the Thesis	5
1.3	Thesis Outline	6
2	Method and Material	7
2.1	Research Approach	7
2.2	Research Design	7
2.3	Data Collection and Analysis	8
3	Existing Knowledge and Best Practice on Building the Data Governance Model	10
3.1	Data Governance Models	10
3.1.1	Data as a Service (DaaS)	11
3.1.2	People	13
3.1.3	Processes	15
3.1.4	Technology	17
3.1.5	Data governance maturity model	21
3.2	Conceptual Framework for Building the data Governance Model	23
4	Current State Analysis of Data Governance in the Case Company	24
4.1	Analysis of the Current Data Governance Model	24
4.2	Key Findings from the Current State Analysis	25
5	Building the Proposal for the Data Governance Model	26
5.1	Introduction to the Proposal Building	26
5.2	Data as a Service (DaaS)	27
5.3	Data Governance Roles and Responsibilities	28
5.3.1	Data owners	28
5.3.2	Data stewards	30
5.3.3	Data engineers	32
5.3.4	Data consumers	33
5.3.5	Organization	34
5.4	Data Governance Processes	35
5.4.1	Create new dataset process	36
5.4.2	Change in access rights process	36

5.4.3	Change in data definitions process	37
5.5	Technology and Architecture	38
5.5.1	Enterprise data management architecture	38
5.5.2	Databases, sources, modelling and integration	39
5.5.3	Metadata, data profiling, quality and business glossary	40
5.5.4	Master and reference data	41
5.5.5	Data storage, analytics and reporting	42
5.5.6	Data catalogue and data portal	42
5.6	Summary of the Proposal	43
6	Validation of the Proposal	44
6.1	Validation Overview	44
6.2	Final Proposal	45
6.2.1	Data as a service (DaaS)	45
6.2.2	Define data governance roles and responsibilities	46
6.2.3	Data owners	46
6.2.4	Data stewards	46
6.2.5	Data engineers	46
6.2.6	Data consumers	46
6.2.7	Organization	47
6.2.8	Data governance processes	47
6.2.9	Technology and architecture	47
7	Discussion and Conclusions	49
7.1	Executive Summary	49
7.2	Practical Implications and Next Steps	51
7.3	Self-evaluation of the Thesis and Final Words	52
	References	54

## List of Figures

Figure 2-1 Inductive Research Approach, McCombes 2020	7
Figure 2-2 Research design	8
Figure 2-3 Data plan	9
Figure 3-1 Four pillars for data governance (Sarkar 2015)	11
Figure 3-2 Data Service Bus (Sarkar 2015)	12
Figure 3-3 Data Services key activities and deliverables (Sarkar 2015)	13
Figure 3-4 Data governance pyramid (Bhansali 2013)	14
Figure 3-5 Enterprise Data Management Reference Architecture (EDM) (adapted from Soares 2014, p.28)	18
Figure 3-6 EDM data and technology	19
Figure 3-7 EDM data management	20
Figure 3-8 EDM analytics	20
Figure 3-9 EDM parallel (matrix) processes	21
Figure 3-10 Gartner data governance maturity model	22
Figure 3-11 Conceptual framework	23
Figure 4-1 Example for user access rights management process	24
Figure 5-1 Data as a Service Framework	27
Figure 5-2 Gartner Enterprise Information Management Framework	28
Figure 5-3 Items for the Data Governance Board, named here information governance (IG)	30
Figure 5-4 Data pipeline	33
Figure 5-5 Different Tiers of BI and Analytics Platforms	34
Figure 5-6 Organizing data projects	35
Figure 5-7 A new dataset process	36
Figure 5-8 Requesting user access rights process	37
Figure 5-9 Change in the dataset definition process	38
Figure 5-10 An overview of the EDM model, with dataflow	39
Figure 5-11 Incoming dataflow	40
Figure 5-12 Metadata handling	41
Figure 5-13 Loading data warehouse	42
Figure 6-1 Excluded organizational parts	47
Figure 7-1 Conceptual framework	50

**List of Tables**

Table 3-1 Differences in data governance models (Gartner webinar)	10
Table 5-1 Data owner role sheet	29
Table 5-2 Data steward role sheet	31
Table 5-3 Data engineer role sheet	32
Table 6-1 Model validation questions and answers	45

## 1 Introduction

What is *data*? There are several definitions of what data is. Most agree on data being information content on the lowest level. It can mean a single cell in a spreadsheet or a single letter in a text document. Data is said to be transformed into *information content* when seen in context or analyzed.

It has also been said that 'data is the new oil' (The Economist, 2017), as the value of companies operating in 'data-business' is much higher than oil companies. Alphabet/Google, Amazon, Apple, Facebook and Microsoft are some of the most valued and most profitable companies in the world. But this has also been challenged (Martinez, 2019), as data is not commonly something that can be sold for profit to external buyers. Data itself has value to the company, but only for specific purposes. There are no 'data-traders' selling data futures, as there are for oil futures.

Nelson (2018) uses the term 'data-rich, but information poor' to describe the modern enterprise. Data is simple to generate or collect from multiple sources, but without context or analysis, there is no information value to it. This simply means that there needs to be an organized way to control how data is put into context or used in the analysis. This organized way is data governance.

Soares (2014) defines data governance as: 'data governance is the formulation of policy to optimize, secure, and leverage information as an enterprise asset by aligning the objectives of multiple functions.' (p.24)

As Ferguson (2018) lists key requirements for data governance, these are:

1. Create a **strategy** for information management
2. Create the right **organizational structure (people)** to produce and govern data
3. Nominate, standardize, and define **data** to be managed
4. Create the right **processes** to produce, manage, and govern data
5. Define **policies** and policy scope to manage/govern specific data
6. Follow a **methodology** to get your data under control
7. Use **technology** to implement policies, and processes to manage and govern data
8. **Produce and publish** trusted data and services for others to consume.

The classical example for data governance failure is NASA Mars Climate Orbiter from 1999. The case was when the Orbiter crashed into Mars due to having different measurement units used in spacecraft operation. One team used English units, while other teams used metric units (Isbell et al., 1999). Throughout nine-month trip from Earth to Mars, this resulted in significant miscalculation and loss of the spacecraft.

This Thesis deals with similar challenges faced by the case company.

## 1.1 Case Company

The Bank of Finland (BoF) is the national monetary authority and the central bank of Finland. Bank of Finland is also part of the Eurosystem, which administers the world's second-largest currency. The Bank of Finland's core tasks are financial stability and financial statistics, banking operations and currency supply (Bank of Finland, 2020). The current chairman of the board is Governor Olli Rehn. Bank of Finland has approximately 360 employees.

ICT and Information Management department is responsible for IT services and tools, as well as document and library services. ICT and Information Management department operates in two locations in Finland, at downtown Helsinki and Vantaa.

The Bank of Finland has an ongoing Data and Analytics development program headed by ICT and Information Management department. This thesis is related to that program.

## 1.2 Business Challenge, Objective and Outcome of the Thesis

The business challenge for the case organization is getting ownership of data assets. A common problem in an enterprise environment is that data is fragmented and in silos. The challenge is to identify where data is and in what format; who owns the data, and what are the legal limitations in using it; as well as how to easily access the data across the enterprise. As the data is used for analysis and forecasting, better access to data improves analysis quality and coverage.



Accordingly, the objective of this thesis is to propose a new data governance model for the Bank of Finland.

The outcome of this thesis is a proposal for a new data governance model for the Bank of Finland.

The scope of the study includes **roles and responsibilities (people), processes, policies, methodology (framework) and technologies** used for data governance. Actual technical solutions, architectures, and physical data structures are excluded, as are data strategy and data definitions.

### 1.3 Thesis Outline

The thesis is organized as follows. The first part is an introduction to topic, business challenge, and case company. This part also contains the objective, scoping and expected outcome.

The second part introduces the research approach and design. Next, a literature study is done for finding modern data governance models. Training and seminars are attended when appropriate.

The third part is describing the current data governance model. Information about current data governance model is collected in discussions with stakeholders. These discussions are facilitated under Data and Analytics development program.

The fourth part is the theoretical context for modern data governance. This is broken down to three topics, people, processes and technical requirements. The fifth part is introducing the proposed data governance model.

The sixth part is the validation of the proposed data governance model. The seventh part is discussions and conclusions.

## 2 Method and Material

### 2.1 Research Approach

Figure 1 illustrates the inductive research approach process.

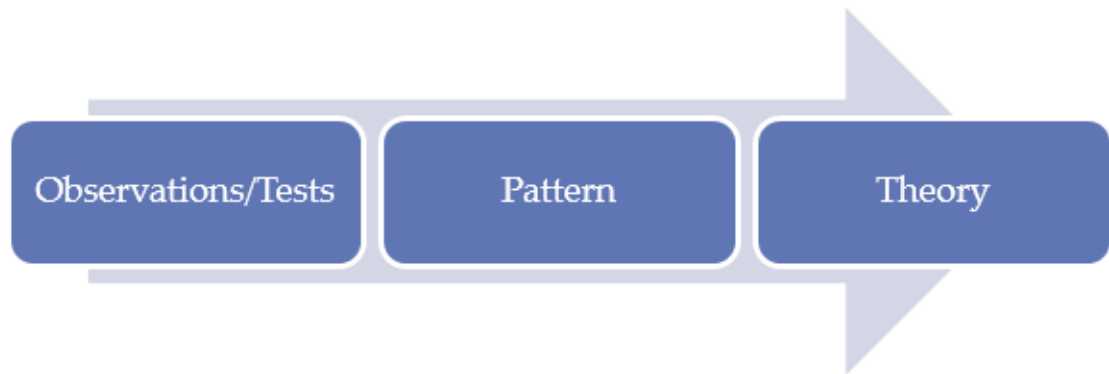


Figure 2-1. Inductive Research Approach (McCombes 2020).

In this thesis, the inductive research approach is used to outline a data governance model based on observations and found patterns.

### 2.2 Research Design

Figure 2-2 shows the research design and related data points in this thesis.

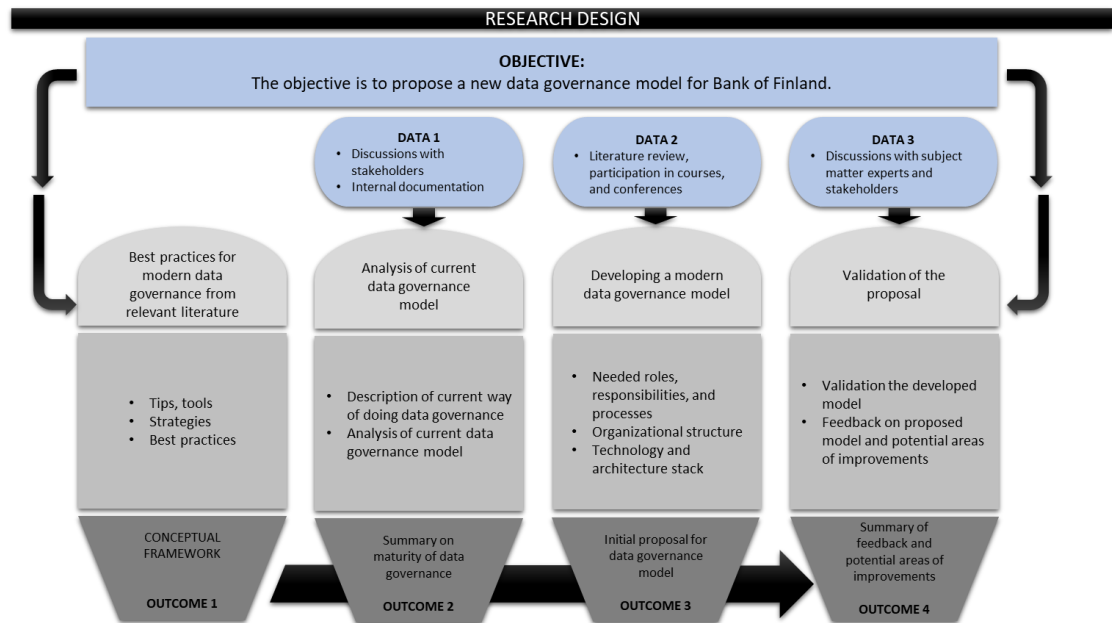


Figure 2-2. Research design.

Both qualitative and quantitative research approaches are used in this study. Some topics are subject to only one or some to both approaches. Best practices for data governance are explored. These are collected from relevant books, publications, and courses. The data governance model is then proposed. Model is then discussed with subject matter experts for feedback. The data is used to build the outcomes.

### 2.3 Data Collection and Analysis

Figure 2-3 shows the data collection that is used to build the intermediate outcomes and create the final proposal.

DATA PLAN					
	CONTENT	SOURCE	INFORMANT	TIMING	OUTCOME
<b>DATA 1</b> Analysis of current data governance model	<ul style="list-style-type: none"> <li>Discussions with stakeholders</li> <li>Internal documentation</li> </ul>	<ul style="list-style-type: none"> <li>Discussions with Stakeholders, one-to-one, and in groups</li> <li>Internal documentation</li> </ul>	<ul style="list-style-type: none"> <li>Data scientist</li> <li>Head of office</li> <li>Head of department</li> <li>Developer</li> </ul>	<ul style="list-style-type: none"> <li>January-February 2020</li> </ul>	<ul style="list-style-type: none"> <li>Summary on maturity of data governance</li> </ul>
<b>Data 2</b> Developing a modern data governance model	<ul style="list-style-type: none"> <li>Needed roles, responsibilities, and processes</li> <li>Organizational structure</li> <li>Technology and architecture stack</li> </ul>	<ul style="list-style-type: none"> <li>Literature review</li> <li>Courses</li> <li>Conferences</li> </ul>	<ul style="list-style-type: none"> <li>N/A</li> </ul>	<ul style="list-style-type: none"> <li>February – July 2020</li> </ul>	<ul style="list-style-type: none"> <li>Initial proposal for data governance model</li> </ul>
<b>DATA 3</b> Validation of the proposal	<ul style="list-style-type: none"> <li>Validation the developed model</li> <li>Feedback on proposed model and potential areas of improvements</li> </ul>	<ul style="list-style-type: none"> <li>Discussions with subject matter experts and stakeholders</li> </ul>	<ul style="list-style-type: none"> <li>Program manager</li> </ul>	<ul style="list-style-type: none"> <li>July-October 2020</li> </ul>	<ul style="list-style-type: none"> <li>Summary of feedback and potential areas of improvements</li> </ul>

Figure 2-3. Data plan.

Research data is collected in formal and informal discussions that were used to define the current state analysis and conduct the validation in this study.

### 3 Existing Knowledge and Best Practice on Building the Data Governance Model

#### 3.1 Data Governance Models

Garner (Judah 2018) defines differences between classic and modern data governance as:

Classic data governance	Modern information and analytics governance
<ul style="list-style-type: none"> <li>• often starts in IT</li> <li>• the general approach, irrespective of use case</li> <li>• compliance focus</li> <li>• seeking for truth</li> <li>• rather defensive</li> </ul>	<ul style="list-style-type: none"> <li>• often starts in the business</li> <li>• the specific approach to different analytics use cases</li> <li>• focus on business needs</li> <li>• building trust</li> <li>• rather an opportunity oriented</li> </ul>

Table 3-1. Differences in data governance models (Gartner webinar, Judah 2018).

Table 3-1 shows the different approaches between classical data governance and modern data governance. The major difference is that in the classical model, IT is the driver in data governance whereas in modern approach, data governance is business first. Another major difference is that the classical model has compliance focus, whereas the modern model focuses on business needs.

Soares (2014) lists three pillars for data governance as people, process and technology. His further ideas specify data governance into four elements, as shown in Figure 3-1.

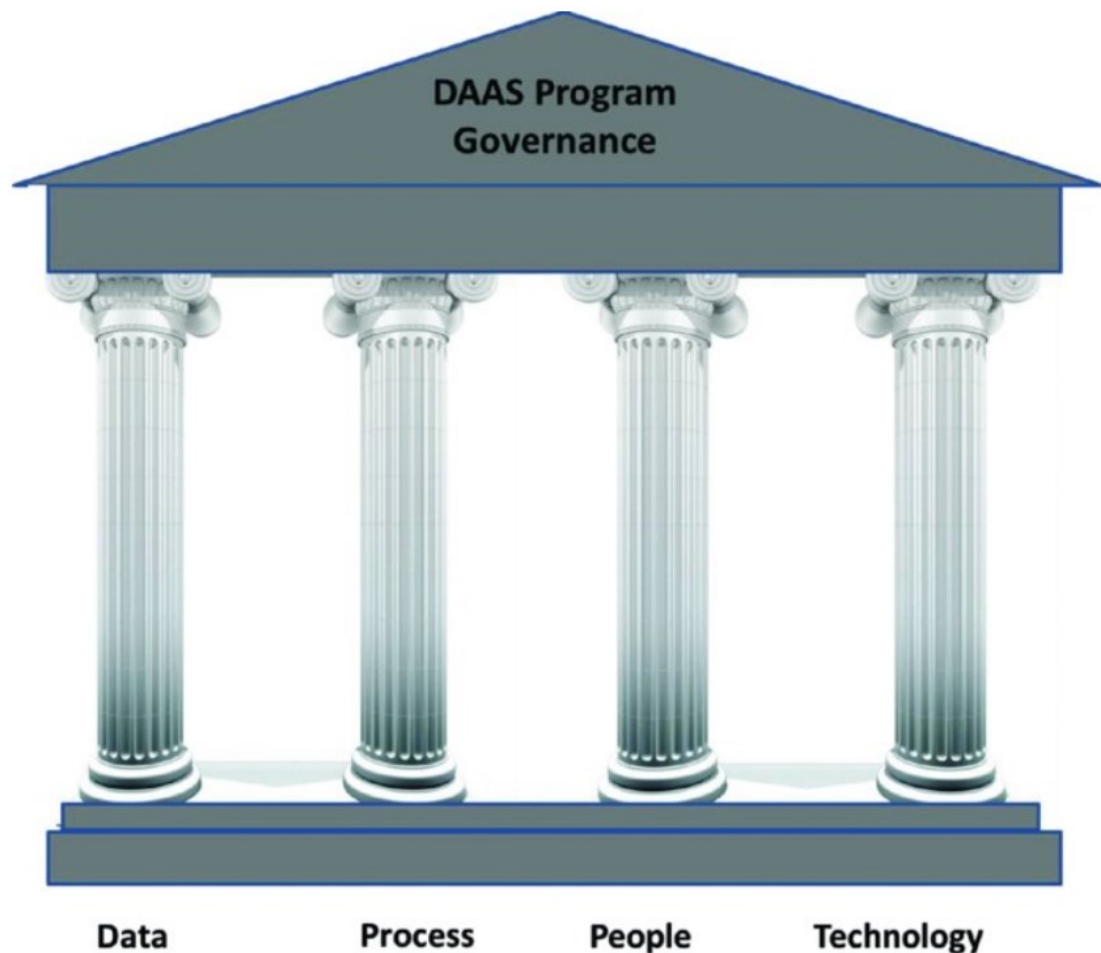


Figure 3-1. Four pillars for data governance (Sarkar 2015).

Sarkar (2015) lists four pillars, data, people, process and technology. These will be explored in the next chapters.

### 3.1.1 Data as a Service (DaaS)

X as a Service (XaaS) model is seen as a modern digital transformation. Newman (2017) writes 'Anyone can choose the services they want, with little to no technical skills or knowledge.' It is seen as a cost-effective tool for creating flexibility and agility.

Data as a Service (DaaS), is seen as a 'marketplace', where data can be accessed. Amazon Web Services launched one to 'include a vast array of curated data sets that are centrally stored, searchable, and managed on Amazon's cloud.' (Linthicum 2016). The idea behind this is to provide easy centralized access to all relevant data from a single location.

Figure 3-2 describes the concept of the data service bus.

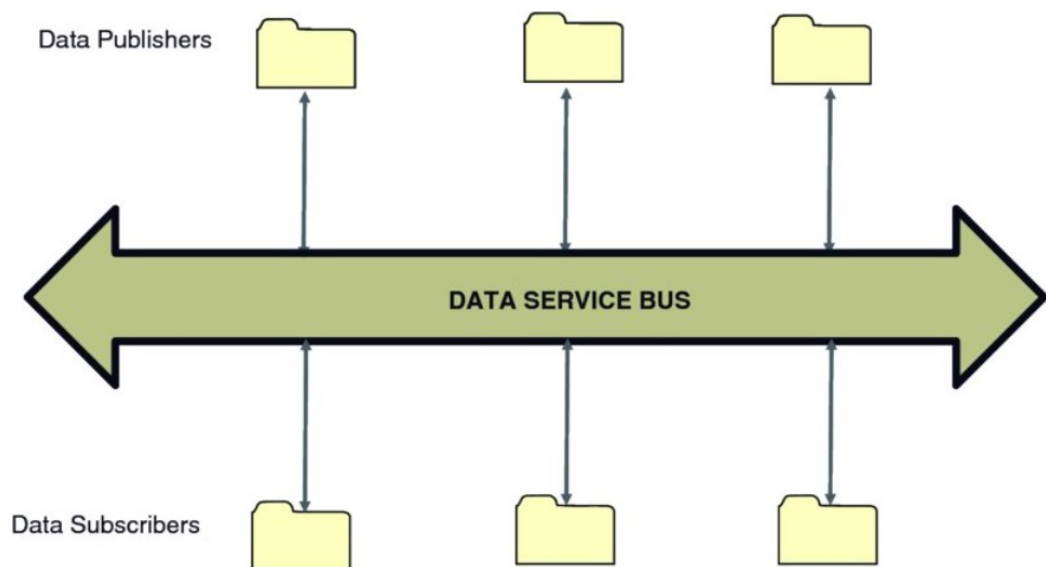


Figure 3-2. Data Service Bus (Sarkar 2015).

Sarkar (2015) uses the term 'Data Service Bus'. The concept behind a data service bus is that it will act as 'foundation for data reuse in any DaaS deployment' (Sarkar 2015). The concept is that there are common data services used in all data operations. It requires common architecture, containing common and reusable data modules. This is seen as a 'data factory', where data is processed in a common 'manufacturing pipeline' that the downstream systems and data consumers use.

Figure 3-3 describes key activities and deliverables in executing data services.

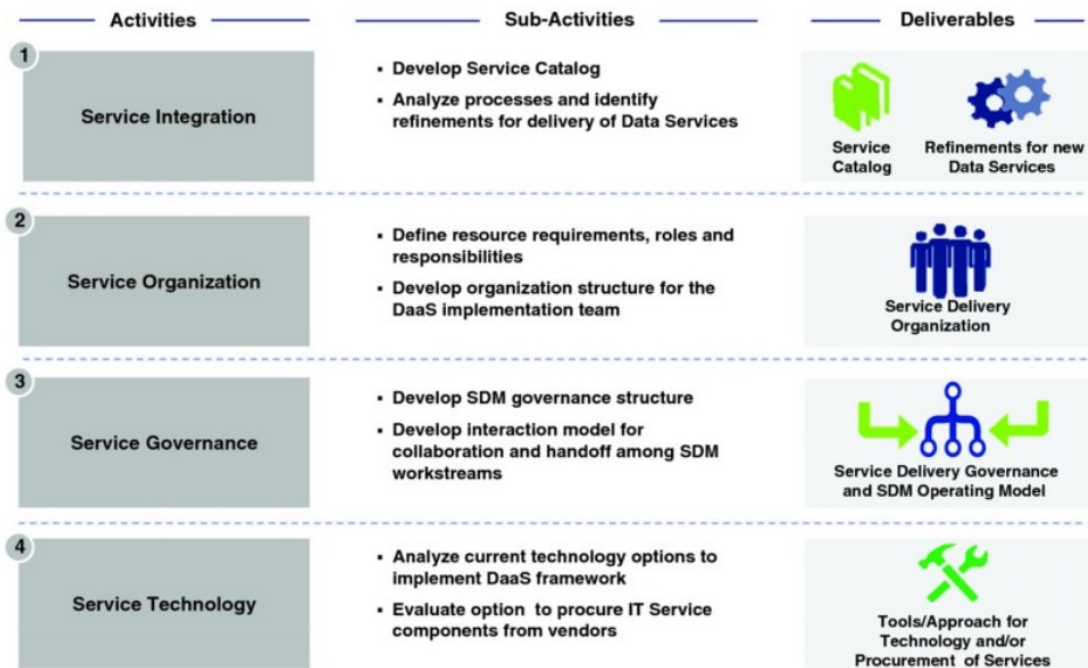


Figure 3-3. Data Services key activities and deliverables (Sarkar 2015).

Figure 3-3 shows the key activities and deliverables in executing data services. These are integration, organization, governance and technology. Here one can see the similarities on building a data governance model.

### 3.1.2 People

Organizational requirements and roles and responsibilities are covered under this topic. Figure 3-4 shows how Bhansali (2013) breaks down the data governance structure to three levels.



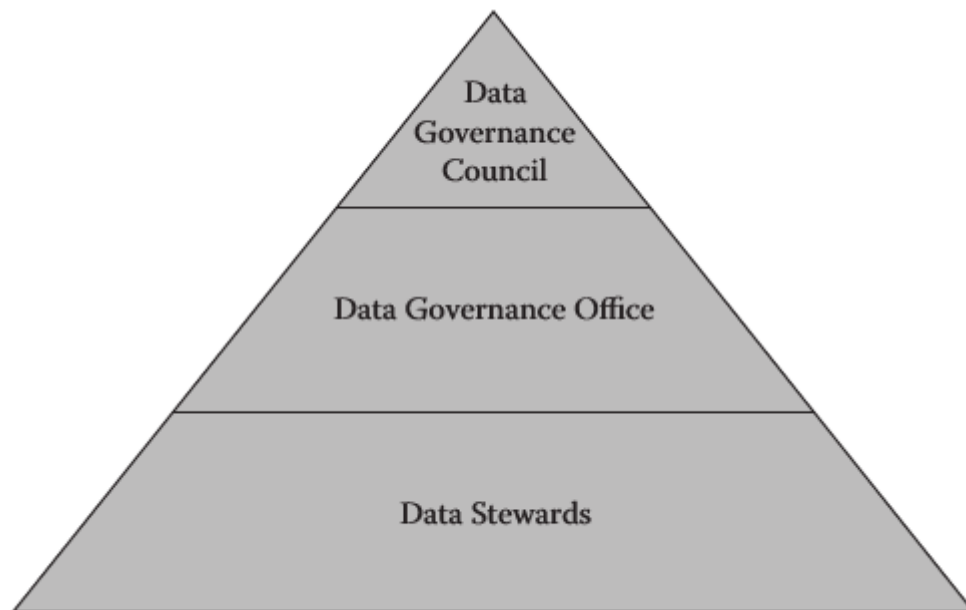


Figure 3-4. Data governance pyramid (Bhansali 2013).

Bhansali (2013) breaks down the data governance structure to three levels, as shown in Figure 3-4. These levels are data governance council, data governance office and data stewards. The highest level is the data governance council. The council or steering group is a 'cross-functional, executive-level group' (Bhansali 2013, p25) which is responsible for policy and strategic decisions. It should include representation from all business and technical stakeholders. These stakeholders are the data owners and approve rules, access and usage policies.

The council is responsible for (adapted from Bhansali 2013, p31): Ensuring that the organization has a data governance strategy

- Balancing the perspectives of stakeholders, users, and IT
- Navigating the organization to ensure data governance
- Approving processes for budgeting, acquiring, and implementing applications and infrastructure
- Approving and modifying the responsibilities of IT and users
- Ensuring that IT applications and activities conform to relevant policies, procedures, regulations, and internal controls.

Second highest level is the data governance office. Data governance office coordinates data governance (strategic) and stewardship (tactical) activities (Bhansali 2013, p25).

Data governance office manages daily activities, communication with stakeholders, project scoping, compliance, and other data activities. Data governance office creates disaster recovery plans.

The lowest level is data stewards. Data stewards include business data stewards, and IT data stewards (data engineers). Data stewards and engineers create processes, handle master and reference data, and operate daily data processes (Bhansali 2013). Each data stewards are responsible for their business area. Their responsibilities include business rules, providing definitions, data quality, and ensuring compliance. IT data engineers' responsibilities include data management, databases, data security and access management.

Bhansali (2013) states that the organization is responsible for successful data governance, which includes:

- Developing and managing the data governance plan
- Developing data standards
- Defining procedures to assess sourcing options
- Managing the portfolio of applications, infrastructure, and services
- Establishing communication mechanisms
- Maintaining relationships with stakeholders.

Data consumers or end-users have responsibilities for (Bhansali 2013, p32):

- Understanding the data activities that support their function
- Ensuring that the goals of data initiatives reflect the function's needs
- Developing specifications for data-related governance and IT projects
- Providing feedback to data stewards on implementation issues,
- application enhancements, and data needs
- Ensuring that data-related applications function properly
- Participating in developing the data governance agenda and priorities.

### 3.1.3 Processes

To publish reliable data to consumers, one needs to agree on several characteristics and have them commonly available (adapted from Sarkar 2015):

1. definition, structure and semantics

2. data quality and integrity
3. changes to the service and the underlying information.

Definition, structure and semantics need to be agreed as a common language. Data quality and integrity is a major topic to ensure reliable data usage. Change management is key to building an agile data governance process. Details for these three bullets are explored next.

Keith (2007) states that data definition should contain:

- a name or label
- a significance statements
- formats
- valid value lists or validation criteria
- valid operations
- ownership details
- usage details
- source
- comments
- configuration information.

According to Keith (2007), naming conventions need to be commonly agreed and easily understood. Significance statements need to include details like information classification. Formats need to include information about data formats, like database, spreadsheet etc. Data validation value or criteria should include details like data type validation, code or cross-reference validation or structure validation. These are to ensure data quality. Valid operations should clearly state, on what the data can be used on, or if combined with other data, what is the classification level. Data ownership is to be clearly stated. Source and possible copyrights need to be included. Comments and configuration information can be used on tracing data during transformation, processing or version control. There needs to be a common process on data definitions, structure and semantics.

According to Keith (2007), data quality-related issues include:

- transaction rework costs
- costs incurred in implementing new systems
- delays in delivering data to decision-makers.

Transaction rework costs refer to a situation, where data needs to be reloaded, to manually to correct errors. This is a common problem with complex data systems. Costs for implementing new systems increases, when data quality is poor, or not well known. Common issues are; not well-formatted data, and complex data operations. These often result in delays in delivering needed data. Ability to track data lineage is necessary for IT to discover potential issues and version conflicts. There needs to be a common process for ensuring data quality.

Data governance change management is introduced next. Change management comes to question in several scenarios. These might be:

- change in ownership
- change in access rights
- change in data definitions
- other change processes.

Handling changes is important in any business process. In data governance, there is a need for a common process for change.

#### 3.1.4 Technology

Soares (2014) uses Enterprise Data Management (EDM) reference architecture as the technology pillar. Figure 3-5 shows EDM reference architecture.

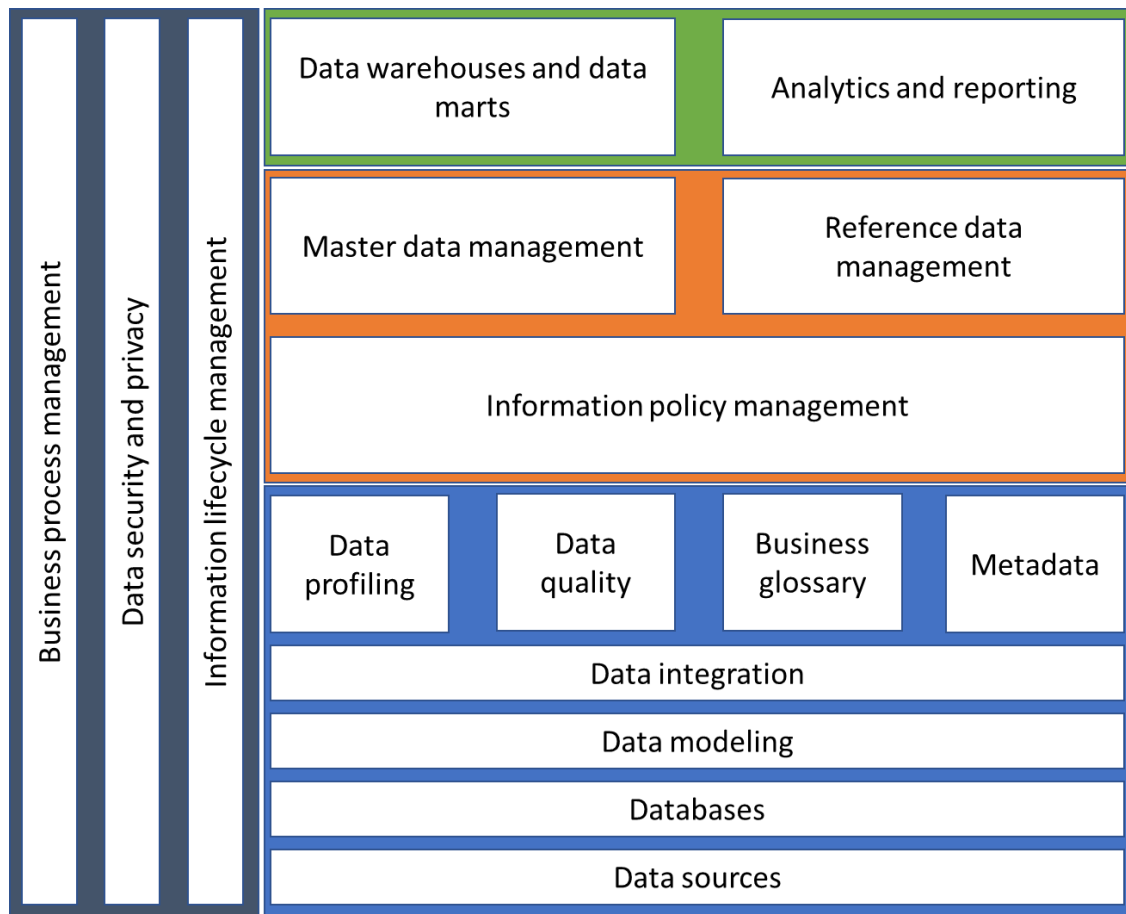


Figure 3-5. Enterprise Data Management Reference Architecture (EDM) (adapted from Soares 2014, p28).

Figure 3-5 shows EDM reference architecture. On the lowest level are data sources that feed into databases. Databases are built using data modelling and provide access to data integration. Above integration, there is data profiling, data quality, business glossary and metadata. These are controlled by information policy management. Above information policy management are master data management and reference data management. The highest level is data warehouses and data marts, and analytics and reporting. Parallel to all is business process management, data security and privacy, and information lifecycle management.

Next, all levels are broken down to four separate parts and examined in detail. Figure 3-6 shows EDM data and technology.

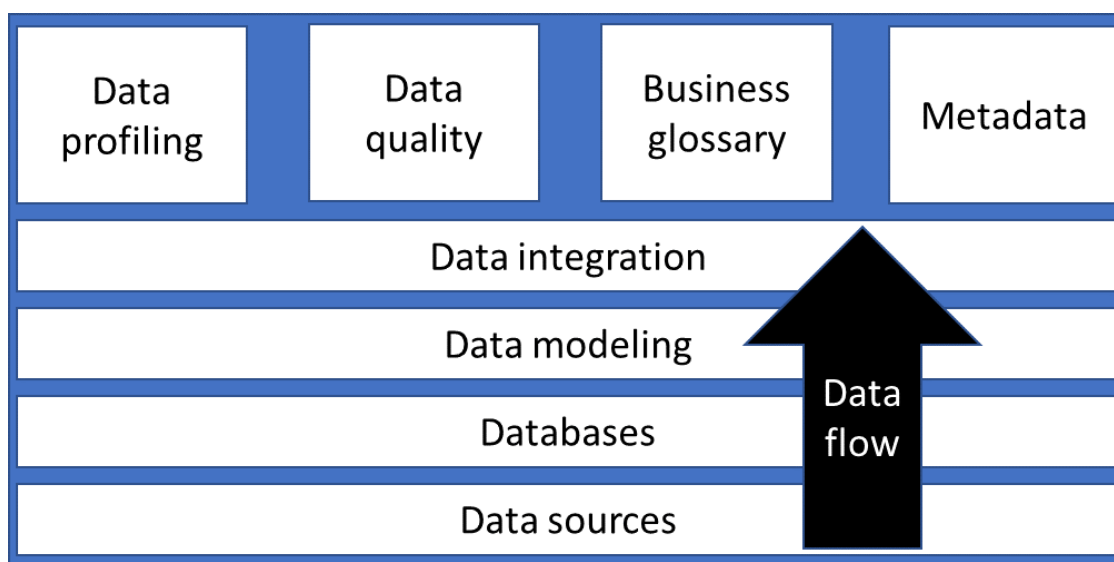


Figure 3-6. EDM data and technology (adapted from Soares 2014).

Figure 3-6 shows EDM data and technology. Data sources refer to incoming data from multiple sources. These can be files, streaming data, social media or data from data brokers. These can be internal or external sources. Incoming data is collected into databases. These can be file-based data lake solutions, or structural data SQL databases, depending on data sources. At this point, data is 'raw-data' and generally not user friendly. Data modelling is the next step to bring structure to incoming data. Data modelling is defining structure to data. It has three levels, conceptual, logical and physical. Conceptual defines the content, logical gives structure and physical level explains actual physical database table structure. To access the modelled data, data is integrated, meaning it can be accessed via tools. Data flow or details for processing the incoming data are not explored in detail in this paper.

To access the data and understand its content, some tools can be used. Data profiling is a process of understanding the data and how it relates to other datasets. This might include statistical analysis. Data quality is the process of understanding and improving the quality and integrity of data. Data quality process often is related to data profiling, which might discover issues with data quality. Business glossary is a set of common key terms which are to be used in the classification of data. Metadata is data of data. It contains details of data artefacts, like name, physical location, quality and relations to other data. These are not explored in detail in this paper.

Figure 3-7 shows the data management part of EDM.

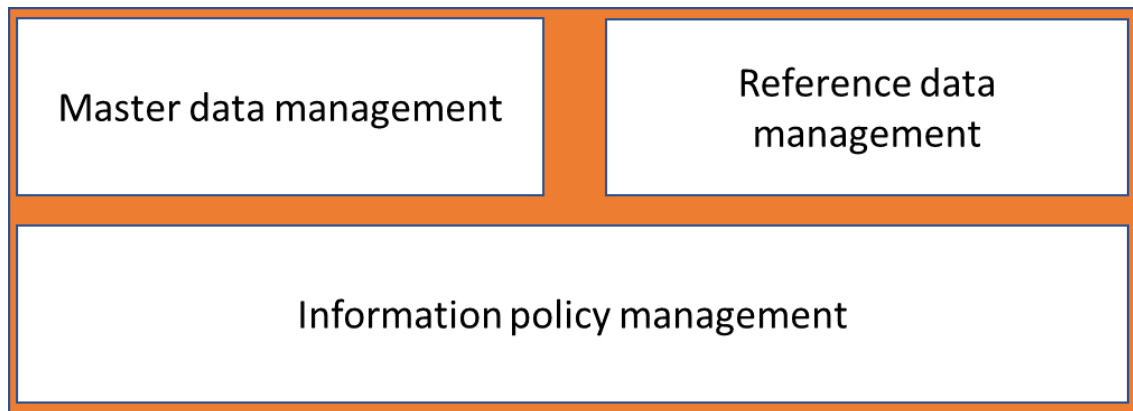


Figure 3-7. EDM data management (adapted from Soares 2014).

Figure 3-7 shows the data management part of EDM. These are information policy management, master data management and reference data management. Information policy management includes governance, risk and compliance details. Master and reference data are single primary versions of common business-critical data. These are things like country and currency codes.

Figure 3-8 shows the EDM analytics.



Figure 3-8. EDM analytics (adapted from Soares 2014).

As shown in Figure 3-7, data warehouses and data marts are data storages for analytics data. Analytics and reporting refer to e.g. Tabular and OLAP cubes, and reporting tools. End-users commonly have access to these datasets.

Next, Figure 3-9 describes processes running over the whole chart.



Figure 3-9. EDM parallel (matrix) processes.

Figure 3-9 describes processes running over the whole chart. Those processes are business process management, data security and privacy and information lifecycle management. Business process management is seen as part of the technical solution for data governance. Data security and privacy are to be considered at all steps in the technical solution. Information lifecycle is an important part of data governance – as data has a lifecycle, after which it is seen as obsolete.

### 3.1.5 Data governance maturity model

Gartner (Taylor 2008) has a data governance maturity model which is used here to evaluate the current status of data governance at the Bank of Finland.

The model has five goals:

1. Data integration across the entire IT portfolio
2. Unification of content throughout the organization
3. Integration of master data domains
4. Smooth flow of information across the organization
5. Metadata management and semantic reconciliation.

Gartner (Taylor 2008) uses six levels to measure maturity for each goal, as shown in Figure 3-10.



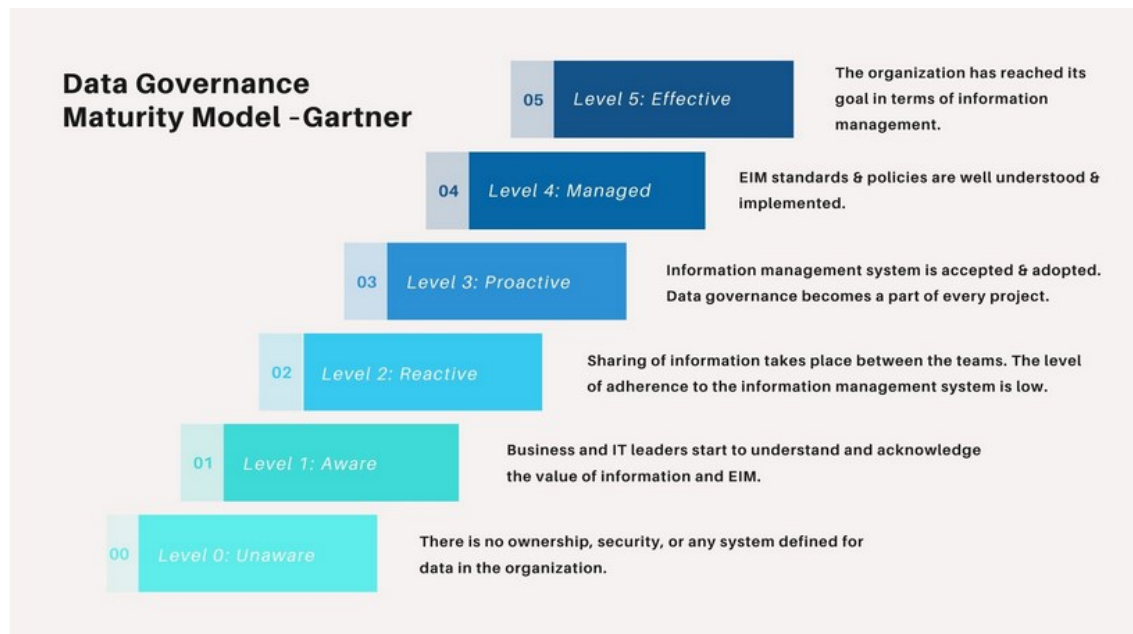


Figure 3-10. Gartner data governance maturity mode Gartner (Taylor 2008).

Figure 3-10 shows the data governance maturity levels. These are in bottom-up order:

- Level 0: Unaware – There is no ownership, the security of any system defined for data in the organization
- Level 1: Aware – Business and IT leaders start to understand and acknowledge the value of information and EIM (Enterprise Information Management)
- Level 2: Reactive – Sharing of information takes place between the teams. The level of adherence to the information management system is low.
- Level 3: Proactive – Information management system is accepted & adopted. Data governance becomes part of every project.
- Level 4: Managed – EIM standards & policies are well understood & implemented
- Level 5: Effective – The organization has reached its goal in terms of information management.

3.2 Conceptual Framework for Building the data Governance Model

Figure 3-11 shows the conceptual framework for data governance.

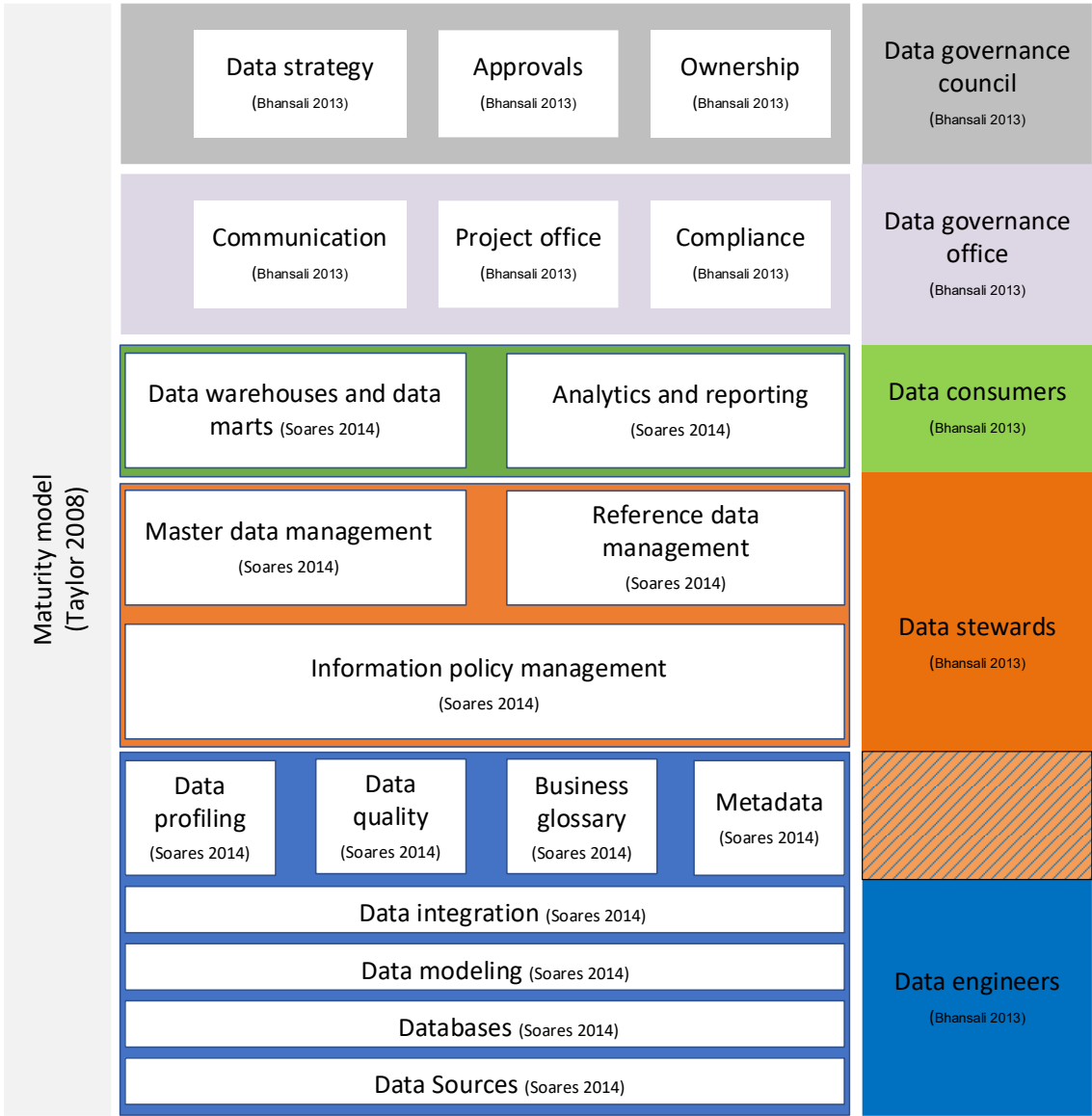


Figure 3-11. Conceptual framework for building the data governance model.

As shown in Figure 3-11, the right side shows the roles and responsibilities. Not all responsibilities are shown in the boxes. The left side shows the previously mentioned reference architecture. As shown in Figure 3-11, there is overlap in data stewards' and data engineers' responsibilities in e.g. data quality and metadata. Both need to maintain their metadata with given datasets. Data governance maturity model used to evaluate the current state is shown on the left side.

## 4 Current State Analysis of Data Governance in the Case Company

The current state analysis is done by using the governance model from Section 3 that was built based on literature to find methods to evaluate the current state. The current state also used the maturity model borrowed from literature. The chosen maturity model is then discussed with stakeholders, to find the correct maturity level. Additional discussions with stakeholders provide insight to current data governance.

### 4.1 Analysis of the Current Data Governance Model

During the discussing with the stakeholders, it was identified that the current data governance in the Bank of Finland is based on data access rights. Data has defined the ownership, and legal status. It is well understood where data is originating, and how it is used. Access rights are generally managed by IT. Data ownership is with business. The following common process is used to give access rights.

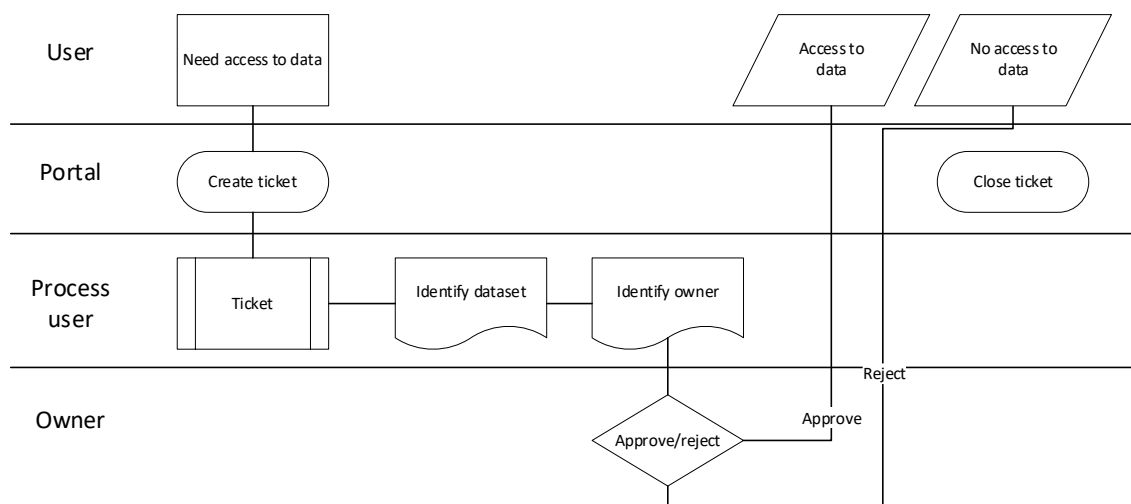


Figure 4-1. Example for user access rights management process.

As shown in Figure 4-1, the current process uses a common IT process tool used at the Bank of Finland. The user creates a ticket in the process tool, and process user identifies the correct dataset and owner. The ticket then goes to the owner for approval. If the owner grants access, process user then manually grants access to the data, usually by adding the user to the appropriate access group.

Reference data is collected and used in different systems, but it is not done uniformly throughout the organization. There are reference databases, and systems, which are

used as sources, but generally all systems copy the reference data into their own databases. Metadata tends to be system-specific and not in a common database.

During the discussing with the stakeholders, it was identified that, generally, the systems operate in organizational silos, and sharing data between systems can be done, but there is no common process or architecture for it.

## 4.2 Key Findings from the Current State Analysis

Key findings from the current state analysis for Bank of Finland include:

1. Co-operation between organizational branches is based on personal contacts, not on commonly agreed processes
2. There are no common processes for a data catalogue; there is no common data catalogue
3. There are no common agreed roles and responsibilities for data engineering
4. There is a common understanding of data ownership
5. There is a common process for granting access to data
6. There is no common agreed architecture for data processing
7. There is no common agreed way of data storage.

Due to the organizational silos, different departments in the Bank of Finland are on different levels. Parts of the organization are – according to the Gartner data governance maturity model – relate to levels 1 and 2:

- Level 1: Aware – Business and IT leaders start to understand and acknowledge the value of information and EIM (Enterprise Information Management)
- Level 2: Reactive – Sharing of information takes place between the teams. The level of adherence to the information management system is low.

Thus, the status for the current data governance matches closely to the classic data governance model by Gartner. As Taylor states, at this level: 'There is a well-recognized need for a standard set of tools, processes, and models in place to establish uniformity across the organization.'

## 5 Building the Proposal for the Data Governance Model

The proposal is built guided by suggestions from literature and best practices, discussions with the stakeholders, and participating in conferences.

### 5.1 Introduction to the Proposal Building

Gartner discusses 'the rules of the game' for data governance. They differentiate the basic difference with the classical and modern data governance as 'compliance as "following someone else's rules," such as a regulator's, but governance is based on the agreement of all stakeholders.' Stakeholders can define 'the rules of the game' to include more than just compliance. These rules should be seen as agile and flexible.

From a management perspective, the key topics are:

1. Ownership, responsibility and accountability need to be clear for data and analytics. The owner of data commonly is not the person who is responsible for the quality of data. Analysts who create new datasets also transfer the responsibility.
2. Delegation of decision rights is a key aspect of data governance. Understanding where responsibility is and who is responsible for any analytics output has to be clear.
3. Successful BI and analytics strategy rely on measuring success. Impact of analysis and key performance indicators define success measures and align data governance program with business objectives.
4. Many big data projects are experimental, with potential value, but with unknown feasibility. Fund projects based on expected business outcomes, and business case. Organizations cannot fund every project, so prioritization is important. Creating an innovation budget for projects with high-risk is recommended.

From an execution perspective, the key topics are:

1. Compliance is not governance, but still important. Regulatory and legal requirements need to be understood.
2. Do not analyze everything just because you can. Expanding the code of conduct to include analytics in the organization.

3. Data and analytics validation processes are needed when sharing data and analytics with a larger community. Document transparency on algorithms and methodologies.
4. Monitor and report compliance and utilization rates. It is important to understand what data and analytics reports are used and when to allow removing the excess.

The following chapter describes the proposed modern data governance model. The proposal is divided into four parts according to the pillars. The first part is data, the second part is process, the third part is people and finally, the fourth part is technology.

## 5.2 Data as a Service (DaaS)

Figure 5-1 shows DaaS framework. Building a complete data as a service model is not in the scope of this paper. DaaS model is reviewed on a high level. Additional studies on the topic are recommended.

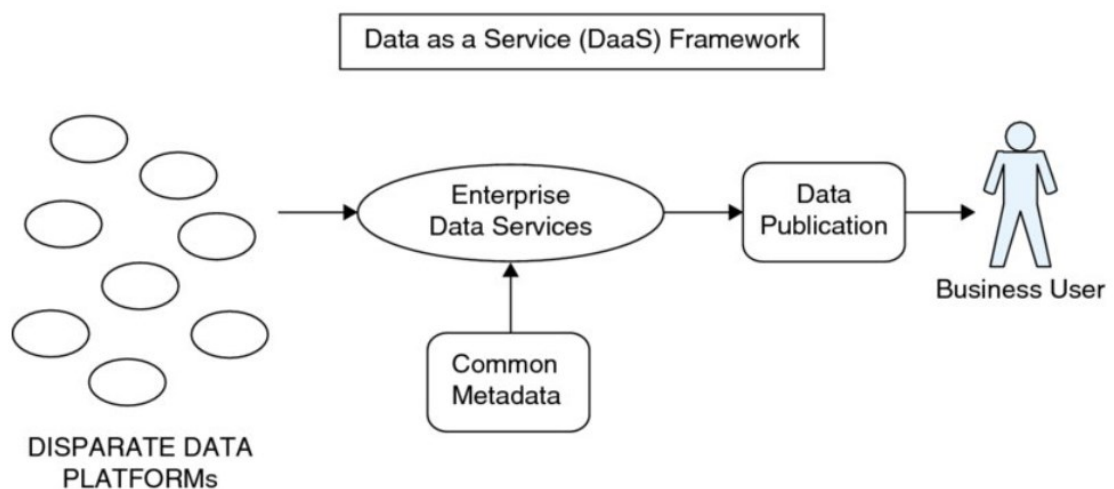


Figure 5-1. Data as a Service Framework.

Figure 5-1 shows DaaS framework. It is a simplified view of how DaaS model should operate. Data from multiple mixed data sources are served to users via Enterprise Data Services. This can be a data virtualization layer or other technology. Idea is to provide all data to users in a common format, standards and definitions. Direct links to data sources are eliminated, and a uniform access provided via Application Programming Interfaces (API).

### 5.3 Data Governance Roles and Responsibilities

Figure 5-2 shows an overview of different roles and responsibilities between business and IT. It defines the roles and responsibilities for data owners, data stewards and data engineers based on the pyramid model. Also, it includes data consumers or users are given roles and responsibilities. Finally, it gives a short description of organizational structure.

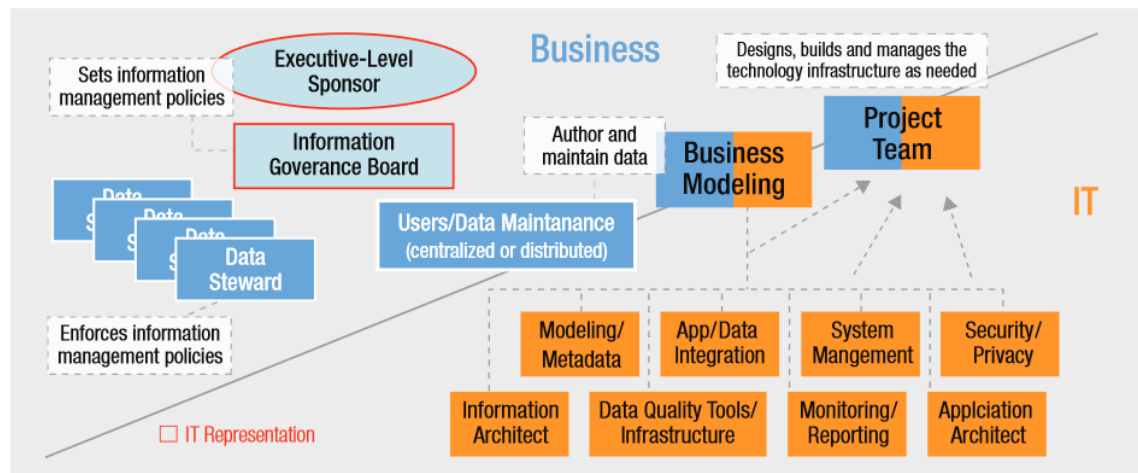


Figure 5-2. Gartner Enterprise Information Management framework.

Figure 5-2 shows an overview of different roles and responsibilities between business and IT. Figure 5-2 also shows the data owners form executive-level sponsorship and data governance council, named the information governance board. It also shows the data stewards form business side operational team. Finally, it shows the data engineers form the IT side operational team.

#### 5.3.1 Data owners

Table 5-1 describes the data owner role. Data owners are the owners of the data. They are to form the data governance council.

Table 5-1. Data owner role sheet.

<b>Job title:</b> Data owner
<b>Job description:</b> A person owning a specified dataset
<b>List of responsibilities:</b> <ul style="list-style-type: none"> <li>• Member of the data governance council</li> <li>• Ensure compliance</li> <li>• Approve data access rights</li> <li>• Approve legal status and classification</li> <li>• Member in data projects steering boards</li> </ul>
<b>Job qualifications and requirements:</b>
<b>Who this role reports to:</b> Data governance council

Table 5-1 describes the data owner role. Most important responsibilities to data owners are to ensure compliance. They also need to review and approve data projects and processes.

Next, Figure 5-3 shows agenda for data governance council meeting.



Item	Description
Current Program Status	Review the most-recent update on the status of any implementation, and roll out in support of the IG work. This could be a system implementation (such as MDM and enterprise metadata management), or a large-scale business application program.
Pending Program Expansion	Leadership reviews major new (business and/or IT) budgeted items that either impact, or are impacted by, IG that will ramp up at some point in the future. This is needed since resources will need to be assigned (or reassigned) in support of the information underpinning the new programs and investments.
Benchmarks, KPIs, and Others	<i>This is the most-important item on the agenda.</i> The board focuses first on business outcome and performance indicators. Then, it drills down into workflow KPIs. Only address specific data quality issues if the information stewards council was unable to address these previously. For data quality issues, at this level, it is an extreme form of exception management. The result might warrant a change in IG policy.
Organization	Discussion of any needs or pending changes in the governance board or information stewards council.
Standards and Policies	Board reviews and approves new and changed IG policies and standards. Senior IT leadership can help guide the business leaders so they understand the impact of their decisions.
Process Improvement	This task is the responsibility of the information stewards, and they should propose to the IG board when, and if, they perceive a need or opportunity to make changes to the work of information stewardship, or the actual business process itself, for example, new product introduction, or new customer onboarding.
Impact Analysis	Business users to determine the financial and business impact on the change in an IG policy, using business simulation. This is the most-sought-after and most-difficult task of all.
IG = information governance; MDM = master data management	

Figure 5-3. Items for the Data Governance Board, named here information governance (IG).

Figure 5-3 shows agenda for data governance council meeting. These include a review of current data program status, pending expansions, current benchmarks, KPIs and other indicators. Data governance council also approves changes to organization, standards and policies. Data governance council will also approve proposals from data stewards, for process improvements, etc. Finally, the data governance council needs to review impact analysis to understand how data programs affect business.

### 5.3.2 Data stewards

Table 5-2 describes the data steward role. Data stewards form the data steward council. The council is responsible for the day to day operations, project management and processes. The council proposes standards and policies for data governance council's approval.

<b>Job title:</b> Data steward
<b>Job description:</b> A business role, the person handling specific dataset
<b>List of responsibilities:</b> <ul style="list-style-type: none"> <li>• Member of data steward council</li> <li>• Enforce data management policies</li> <li>• Define policies, processes and standards</li> <li>• Conflict resolution</li> <li>• Business modelling</li> <li>• The data project manager, or project member</li> <li>• Maintain data</li> </ul>
<b>Job qualifications and requirements:</b> Business role, business process understanding
<b>Who this role reports to:</b> Data governance council

Table 5-2. Data steward role sheet.

Key tasks for the data stewards include:

- Establishing a review and approval process for data definitions, domain-value specifications, and business rule specifications.
- Resolution of conflicting data definitions among multiple stakeholders of that information.
- Establishing information-related policies, standards, and guidelines for compliance across the enterprise.
- Establishing appropriate measures and SLAs to monitor performance improvements in the realm of data- and service-quality efforts.
- Establishing consistent data access because data visibility policies need to be enforced for all data services. There need to be adequate data security controls for all company data.
- The data stewards should also try to keep information open to all employees. For some areas, access does need to be restricted, as the need to keep confidential information safe and secure should be given top priority.

### 5.3.3 Data engineers

Table 5-3 describes the data engineer role.

<b>Job title:</b>
Data engineer
<b>Job description:</b>
IT role, the person handling data pipelines
<b>List of responsibilities:</b>
<ul style="list-style-type: none"> <li>• The data project manager, or project member</li> <li>• Maintain data pipelines; plan, design, operate and troubleshoot</li> <li>• System reliability and performance</li> </ul>
<b>Job qualifications and requirements:</b>
Technical role, understanding technical requirements
<b>Who this role reports to:</b>
Data steward council, data governance council

Table 5-3. Data engineer role sheet.

Data engineer is an IT role. The essential requirements for the role of data engineer are:

- Excellent knowledge of SQL and Python
- Experience with cloud platforms
- Good understanding of SQL and NoSQL databases (data modelling, data warehousing).

Data engineers are responsible for building data pipelines. Figure 5-4 shows an example of a data pipeline.

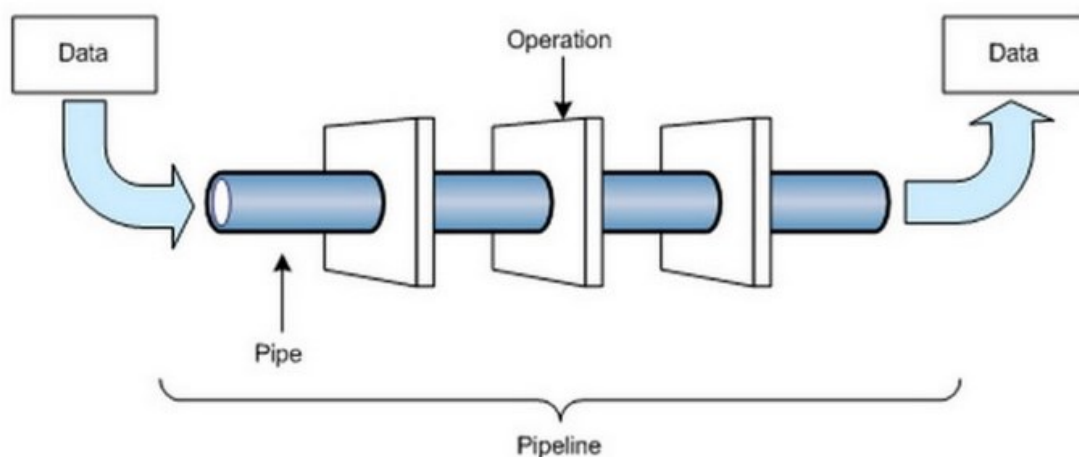


Figure 5-4 Data pipeline

Figure 5-4 shows an example of a data pipeline. Data engineers work with pipelines, build them, supervise them and update when needed.

#### 5.3.4 Data consumers

Data consumers are consuming data. Data consumers are responsible for following set policies, guidelines and standards. There are multiple different types of data consumers. Figure 5-5 shows different tiers in BI and analytics usage.

	Information Portal	Analytics Workbench	Data Science Lab
<b>Key Attributes</b>	<ul style="list-style-type: none"> <li>■ Execution and Efficiency</li> <li>■ Predominantly descriptive analytics</li> <li>■ Represents corporate standard</li> </ul>	<ul style="list-style-type: none"> <li>■ Exploration and visualization</li> <li>■ Predominantly diagnostic analytics</li> <li>■ Continuous and ad hoc analytic processes</li> </ul>	<ul style="list-style-type: none"> <li>■ Experimentation and innovation</li> <li>■ Extends analytics spectrum with predictive and prescriptive analytics</li> <li>■ Addresses new ideas and complex business problems</li> </ul>
<b>Focus</b>	Data models and information governance for analytics	Analytics models and business analyst self-service	Prediction models and data science projects
<b>Key Governance Objectives</b>	<ul style="list-style-type: none"> <li>■ Efficiency</li> <li>■ Integrity</li> <li>■ Confidentiality</li> <li>■ Fidelity</li> <li>■ Quality</li> </ul>	<ul style="list-style-type: none"> <li>■ Alignment</li> <li>■ Relevance</li> <li>■ Consistency</li> </ul>	<ul style="list-style-type: none"> <li>■ Viability</li> <li>■ Transparency</li> <li>■ Privacy</li> <li>■ Legitimacy</li> </ul>

Figure 5-5 Different Tiers of BI and Analytics Platforms

Figure 5-5 shows different tiers in BI and analytics usage. These can be seen as data consumer roles. Users only accessing data through an information portal require limited access rights – read-only. Analytics users build their models and have different needs. Data scientists' requirements are the most detailed and require the most access to data. Different consumer roles have different requirements for data governance.

### 5.3.5 Organization

The proposed roles and responsibilities require changes in organizing data projects. A new role, the Chief Data Office (CDO) is proposed. CDO is the chair for the data governance council, and responsible for all data projects. For data projects, new Data Project Management Office (PMO) is needed.

Figure 5-6 shows how CDO is responsible for PMO operations and DGC chair.

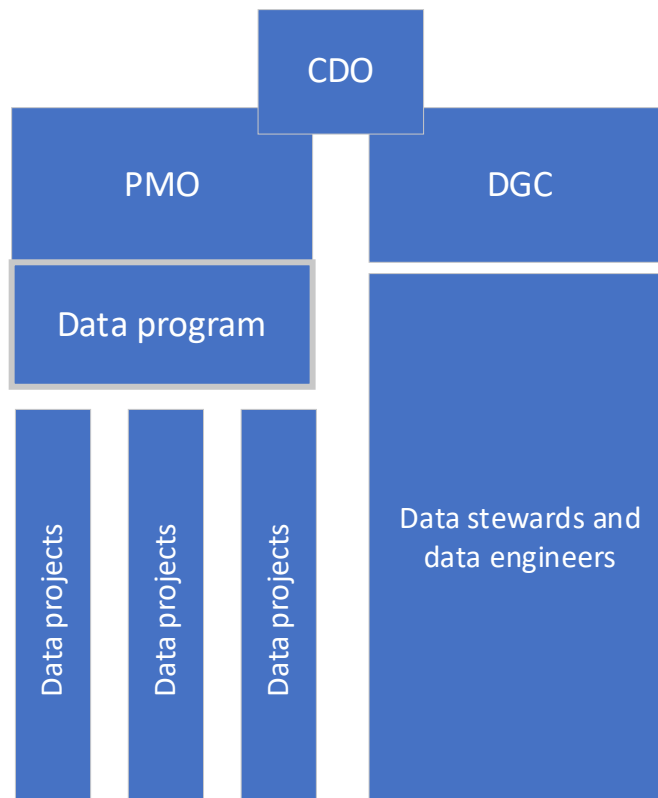


Figure 5-6 Organizing data projects

Figure 5-6 shows how CDO is responsible for PMO operations and DGC chair. Data program is under data PMO and data projects under the program. Data stewards and data engineers participate in data program and projects. PMO is responsible for monitoring and allocating resources to program and projects.

#### 5.4 Data Governance Processes

Data governance processes relate to the governance-related processes. Additional business and IT processes are needed to add to the data governance model but are not covered here. Final processes are created by the data steward council and approved by the data governance council. The following section focuses on change processes.

### 5.4.1 Create new dataset process

‘Create new dataset process’ covers cases when a new dataset is available and ownership is defined. Usually ownership for a dataset is clear and is derived from the ownership of the corresponding business process. Figure 5-7 shows how a new dataset is added.

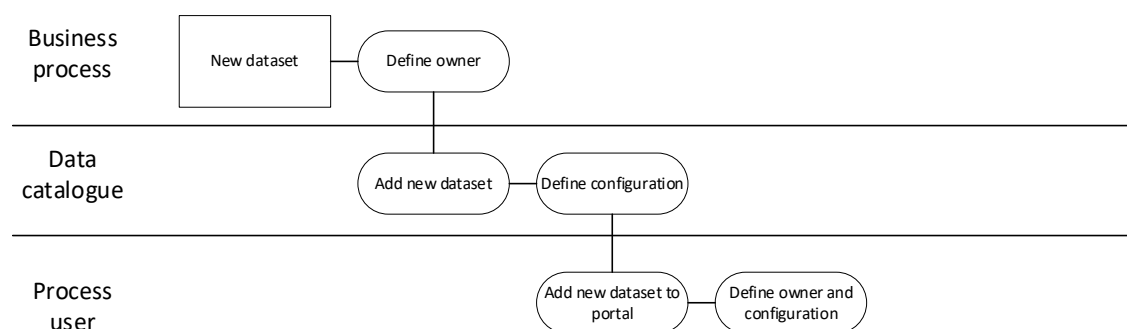


Figure 5-7. A new dataset process.

Figure 5-7 shows how a new dataset is added. The new dataset is created in a business process and ownership defined. The dataset is added to data catalogue, and configuration is defined. The dataset is added to process portal and configured to include information e.g. ownership. Data catalogue and process portal are covered in the technology section in more detail.

### 5.4.2 Change in access rights process

Requesting the additional access rights is a simple process where approval needs to come from the owner. Figure 5-8 shows how user access rights can be requested.

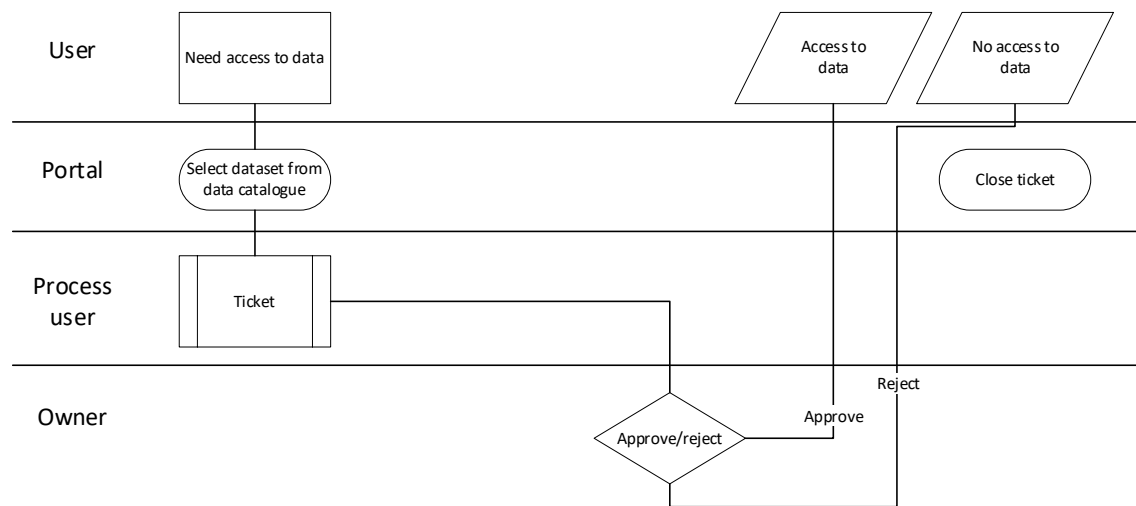


Figure 5-8. Requesting user access rights process.

Figure 5-8 shows how user access rights can be requested. When a user needs access to data, he/she goes to a portal and selects from the data catalogue to which dataset the access is needed. The process can include multiple datasets in one request. A ticket is created and a request is sent to the owner/s. The owner can approve or reject the request. If approved, the user is added to the appropriate access rights group. The portal and process user are described in more detail in the technology section.

#### 5.4.3 Change in data definitions process

Data definitions can include things such as data structure, access rights group or legal status. Data definitions include also information about the data life cycle. Figure 5-9 shows how a change in the dataset definition is done.



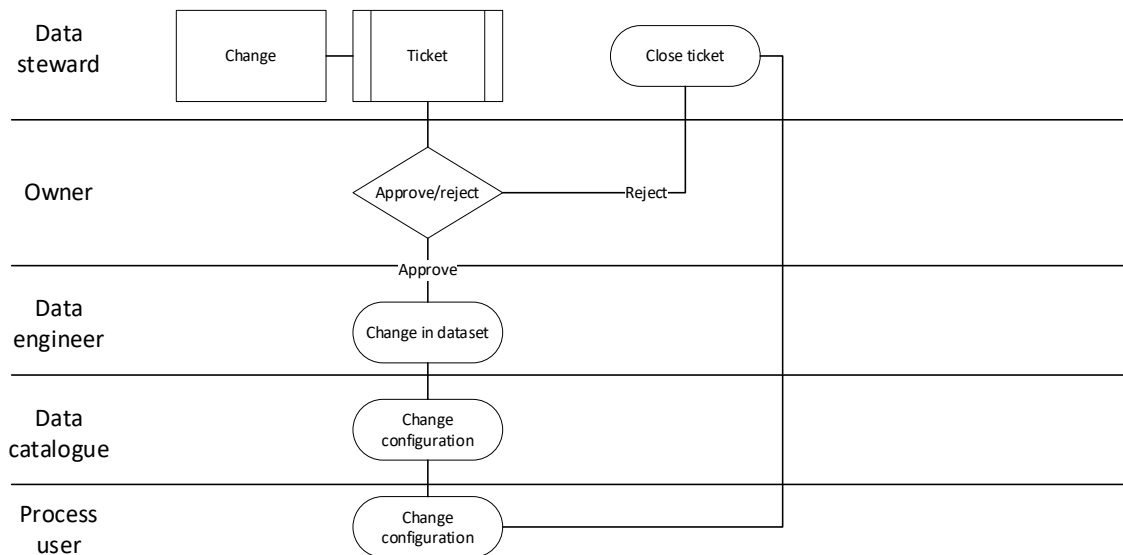


Figure 5-9. Change in the dataset definition process.

Figure 5-9 shows how a change in the dataset definition is done. Data steward identifies a change, creates a ticket and send it to the owner for approval. The owner can approve or reject the change. If approved, data engineer will execute the change in data. Data steward then updates the change in the data catalogue. Data steward can also update the configuration in the portal if authorized, or process user can update the portal.

## 5.5 Technology and Architecture

Data governance technology and architecture include enterprise data management architecture, data modelling and integration, master and reference data management, and data catalogue and data portal.

The purpose here is not to present a complete and detailed technology and architecture description but an introduction to relevant concepts and technologies. For a more detailed picture, an additional study is recommended.

### 5.5.1 Enterprise data management architecture

Figure 5-10 shows an overview of the EDM model.

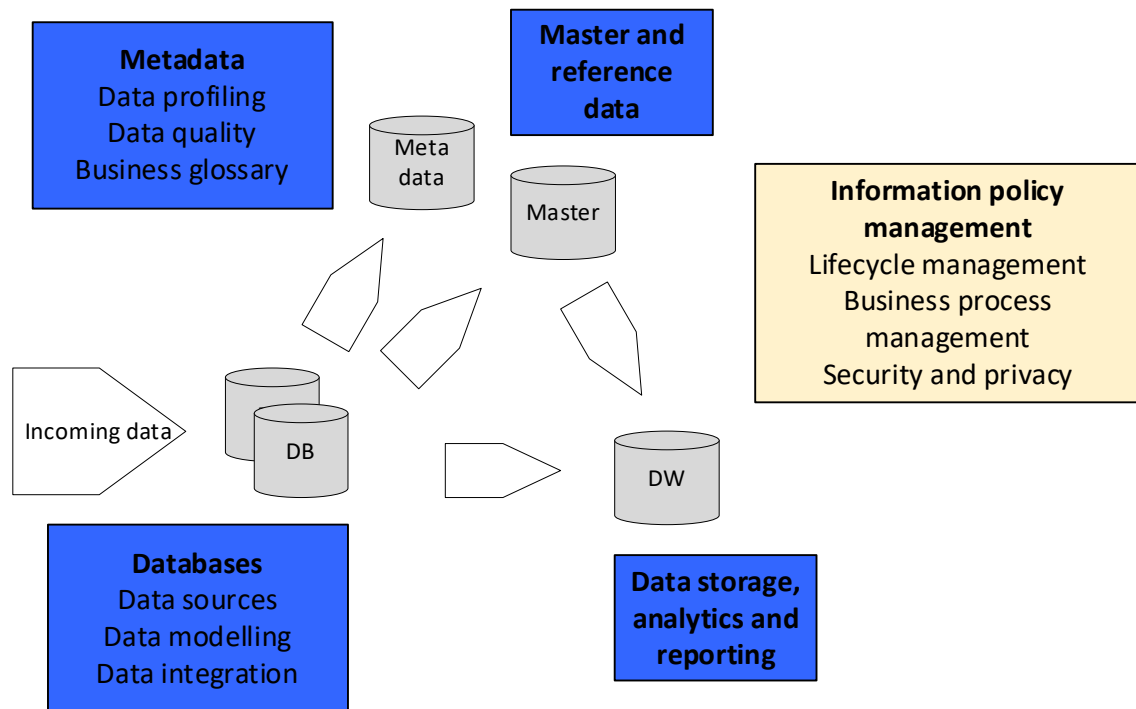


Figure 5-10. An overview of the EDM model, with dataflow.

Figure 5-10 shows an overview of the EDM model. As seen from Figure 5-10, the topics are collected under five common areas. The left side has incoming data loaded into databases. Databases include data modelling and integration. Metadata is extracted from loaded data into a metadata storage. This includes data profiling and quality details, and agreed business glossary.

As seen from Figure 5-10, master and reference data are updated with incoming data if needed. This is done in a separate master data storage. Data is loaded to data warehouses for analytics and reporting needs. The loading can include master and reference data. Information policy management covers data lifecycle management, business processes, and security and privacy policy. Topics are presented in detail in the following chapters.

### 5.5.2 Databases, sources, modelling and integration

Figure 5-11 shows how incoming data is loaded to databases and modelled to fit the purpose.

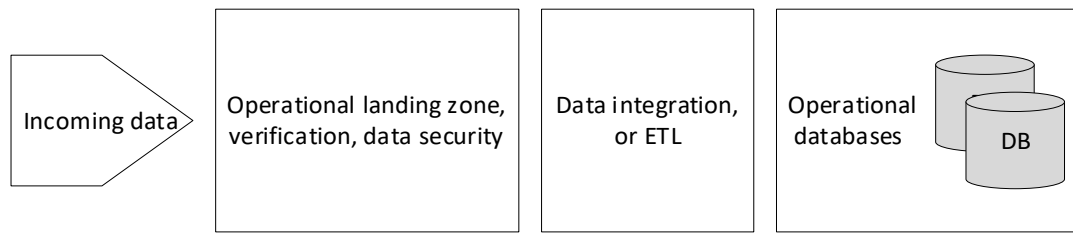


Figure 5-11. Incoming dataflow.

Figure 5-11 shows incoming dataflow process. Incoming data arrives in the business process defined format. This can be e.g. files, streaming data, direct load to the database, or via an API interface. First data needs to be verified to ensure correct, accepted format and to ensure there are no errors and faulty data. Data security is a concern at this point, data needs to be cleansed from any potential hazards, or rejected if any concerns are found.

The business process defines the planned data model. This is a predefined form; which data is to be processed into. There are different levels in data modelling:

1. The conceptual data model describes the semantics of a domain. It covers core concepts, rule and definitions of a business process. The number of objects should be very small.
2. Logical data model shows a detailed representation of data, independent of technology and described in business language.
3. The physical data model is a representation of data design in physical database format.

Data integration or Extract Transform Load (ETL) process is used to format incoming data into the desired form and loaded into databases. Metadata is extracted from incoming data.

### 5.5.3 Metadata, data profiling, quality and business glossary

Figure 5-12 describes metadata, data profiling, data quality and business glossary.

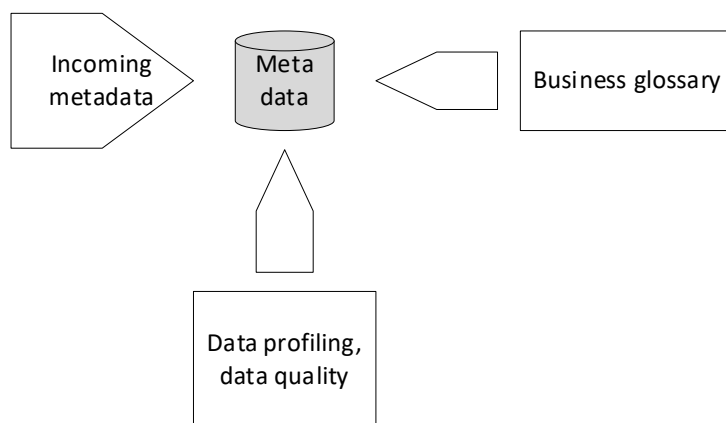


Figure 5-12. Metadata handling.

As seen from Figure 5-12, metadata is data about data. Metadata can contain details like incoming data dates, versions, sender, file format etc. Metadata can be used to track data over its lifecycle. Metadata is extracted from incoming data at the operational landing zone, or integration process. Metadata can be business-related or technical.

Data profiling is used to evaluate data. Profiling includes details about the data, e.g. the number of rows, columns, zeroes, empty cells. This metadata can be used in evaluating data quality. Other data quality evaluations include data content verification, is the data correct. Often it is not possible to evaluate if the data is correct or not. Business processes should define how data quality is evaluated.

Business glossary is a commonly agreed glossary maintained by data stewards. The business glossary can include naming conventions and details about data models. The business glossary also feeds data into operational databases, and integration process.

#### 5.5.4 Master and reference data

Master and reference data are commonly used datasets which are not often updated. When needed, master and reference data is updated with incoming data from operational databases, based on the business process. Master and reference data can include details e.g. country, and currency codes. History and versioning for master and reference data is a must if there are changes over time.

### 5.5.5 Data storage, analytics and reporting

Data storage for reporting and analysis purposes is called a data warehouse. A data warehouse is a long-term data storage. The data warehouse is commonly used for business reporting, and data is modelled based on business needs. Figure 5-13 shows loading data warehouse with data from operational databases, metadata, and master and reference data.

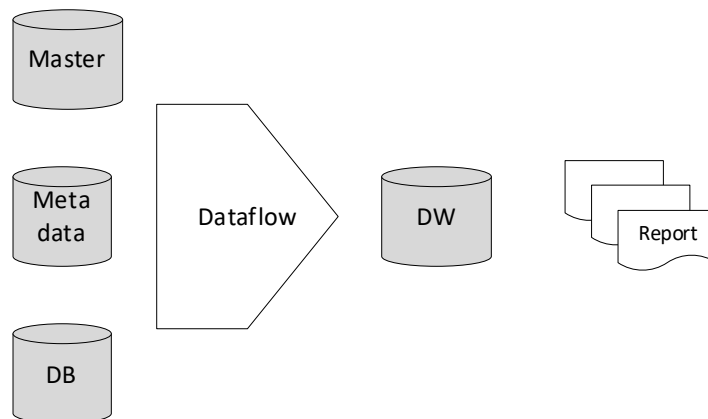


Figure 5-13. Loading data warehouse.

Figure 5-13 shows loading data warehouse with data from operational databases, metadata, and master and reference data. The loading is based on business process needs. A data warehouse should not be directly accessed by data consumers. Reports and reporting databases can be generated from a data warehouse.

### 5.5.6 Data catalogue and data portal

The data catalogue is a collection of metadata, data quality, business glossary and other sources. Purpose of a data catalogue is to enable users to easily find needed data from multiple databases, files, and other sources. Data catalogue should include details about the data integration process, so data can be traced to its source if needed. Data catalogue should handle multiple versions of data.

Data portal in its simple form is an access interface to data catalogue. It should contain details about data, and ability to create e.g. access rights requests, as discussed in 5.3 process section. Data consumer has access to a web interface and can see datasets listed in the data catalogue. If the consumer requires a specific dataset, he/she can select

that dataset and request access rights. A ticket is created in the process tool and sent to the owner for approval. There might be limitations on what the data consumer can see in the data portal. Some datasets might be visible for a limited audience only.

## 5.6 Summary of the Proposal

The proposal contains:

1. Introduction to data as a service model. The model is not explored in detail.
2. Organizational stack, people and roles. These include responsibilities of data owners, data stewards, data engineers and data consumers. Roles are defined and new organizational structure proposed. Formation of data governance council and data steward's council is proposed.
3. Data governance processes are introduced. Change processes are critical to any governance model. Only few processes are reviewed. Development of additional processes is the responsibility of data steward's council.
4. Data governance related technology stack is introduced. These include data management architecture.

## 6 Validation of the Proposal

### 6.1 Validation Overview

The proposal is based on literature review, discussions with the stakeholders and conferences. Material collected from multiple sources is reviewed and the proposal is built. The subject of the study is data governance, but there are no right and wrong answers on if the outcome is correct or not.

Validation was done based on Ferguson's (2018) lists of basic requirements for data governance. According to Ferguson (2018), data governance the management includes:

- Data naming, and data definitions
- Enterprise metadata
- Data modelling
- Data quality
- Data integration
- Data privacy, and access security
- Data retention
- Enterprise content.

To open up these topics, Ferguson (2018) has a list of questions which need to be answered in the data governance model. These questions were used to evaluate the proposed model in this thesis.

Question (Ferguson 2018)	How answered in the proposal in this thesis
What data needs to be controlled?	Controlled data is defined to ensure that only correct persons have access to the data. Data controls are defined by data steward's council.
Where is that data?	Physical data location in databases is documented in the data catalogue. Data engineers know where data is located, and in which format.
What data names is it known by?	Database aliases may be used. Data names are listed in the data catalogue and are known by data engineers.
What should it be known by?	Naming conventions are agreed by data stewards.

What state is the data in, and who is responsible for its quality?	State and quality of data is the responsibility of data stewards.
Does it need to be cleaned, transformed, integrated & shared?	Data ETL operations are executed by a data engineer. Data stewards define the need for cleaning, transformation, integration and sharing.
What transformation has been applied since capture?	A data engineer is responsible for transformation. Data catalogue contains information about ETL.
Should it be synchronized?	A data engineer is responsible for synchronization between databases.
Who is allowed to access, and maintain and are they audited?	Data owners allow access to data and approve maintenance and auditing rules.
How long does the data have to be kept?	Data lifecycle is defined by the data owner.

Table 6-1. Validation questions and answers (based on Ferguson 2018).

Table 6-1 lists the questions which were used to validate the proposed model in this thesis. Left side column has the question and right-side column the answers to the questions. Right side answers cover the responsibility for the topic within the proposed model. All questions are answered based on the proposed model. No question is left unanswered, and all topics have a responsible party.

## 6.2 Final Proposal

Next, the final proposal has been updated based on the discussions with stakeholders. No formal discussions or questionnaires were done; therefore, the proposal is not a complete data governance model. There are several topics which require additional study in building a complete model. This paper is seen as an introduction to the topic. The following sections review the proposal.

### 6.2.1 Data as a service (DaaS)

DaaS model is an important part to a modern data-oriented way of working. It is important to conduct additional study on the topic. Generally moving towards more data-oriented thinking is seen as beneficial.



## 6.2.2 Define data governance roles and responsibilities

In the proposal, roles and responsibilities are defined. Common agreement is to follow up on the proposal. Data owners form the executive-level sponsorship and the data governance council. Data stewards form the business side operational team. Data engineers form the IT side operational team.

### 6.2.3 Data owners

Data ownership is clear. Owners need a common forum to discuss topics, named data governance council. Owners will monitor and form steering boards to data related projects and review effect on business.

### 6.2.4 Data stewards

Key tasks for data stewards need to be clarified. Data stewards are to be nominated and to form data steward council. The council proposes standards and policies for data governance council approval.

### 6.2.5 Data engineers

Requirements for data engineers need to be clarified. At this point, it is not clear how the data engineer role will be filled. Responsibilities in building data pipelines are clear.

### 6.2.6 Data consumers

Data consumers have different knowledge levels and requirements. It is understood that different data consumers will have different needs and requirements. Different level of users has different requirements for data governance.

### 6.2.7 Organization

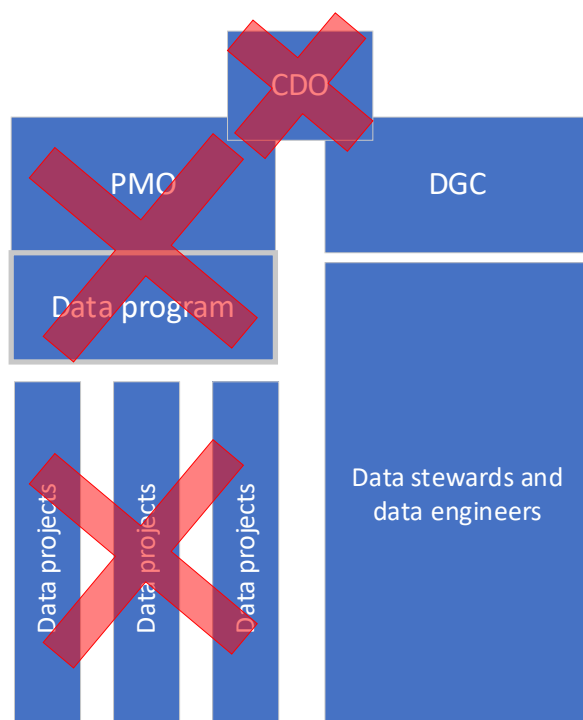


Figure 6-1. Excluded organizational parts.

Biggest challenges were done in re-organizing tasks. In organizational structure CDO role, PMO, and separate data programs were excluded. These are not seen as needed at this point. Current resourcing is limited to the roles and responsibilities that need to be formed in parallel to the current other responsibilities.

### 6.2.8 Data governance processes

Additional processes are needed. It is understood that data steward's council will create and propose processes to be approved by the data governance council.

### 6.2.9 Technology and architecture

The Bank of Finland's IT Architecture group is responsible for technology and architecture. Developing data architecture falls under their mandate. Data architecture should be developed in co-operation with data stewards and engineers. Additional study is needed.

Data modelling and integration is to be taken into account in data related projects. Data modelling should follow the 'conceptual, logical, and physical' design path.

Additional effort needs to go into collecting metadata, data profiling, and data quality topics. Business glossary is needed. Data stewards are responsible for maintaining business glossary.

Finally, the master and reference datasets need to be cataloged and commonly agreed. Data catalogue metadata collection is needed. Additional study is needed to identify the best data catalogue. Data portal needs to be built to ease access to data.

## 7 Discussion and Conclusions

### 7.1 Executive Summary

Data is an important asset to organizations and it needs to be utilized when possible. This requires an agreed and implemented data governance model. Currently, the Bank of Finland has a classical data governance model. The solution proposed in this study is to modernize data governance.

The study started with identifying the business challenge. The challenge was and is how to best utilize data assets and remove the usage barriers. Next, the study was conducted by investigating the current state. No formal questionnaires were used, but the current state analysis was based on discussions with stakeholders. Literature study and participation in conferences were used to find best practices.

A proposal was formed based on the identified best practices. The proposal includes four pillars for data governance. These pillars are roles and responsibilities (people), processes, policies, methodology (framework) and technologies.

First, Data as a Service is seen as a modern way to access data. It provides a centralized access to relevant data assets in a single location. Data factory architecture provides such a common data service model.

Second, the people part includes the organizational structure and roles. Modern data governance includes data engineers, data stewards, data consumers, and data governance council members, who are the data owners.

Third, for data governance process, data definition, structure and semantics need to be agreed as a common language. Data quality and integrity is a major topic to ensure reliable data usage. Change management is key to building an agile data governance process.

Fourth, technology for data governance includes enterprise data management architecture, data modelling and integration, master and reference data management, and data catalogue and data portal.

Summing up, the proposed data governance model follows the conceptual framework developed in this thesis based on suggestions from literature and best practice. The adapted conceptual framework shows the following overview of the proposal.

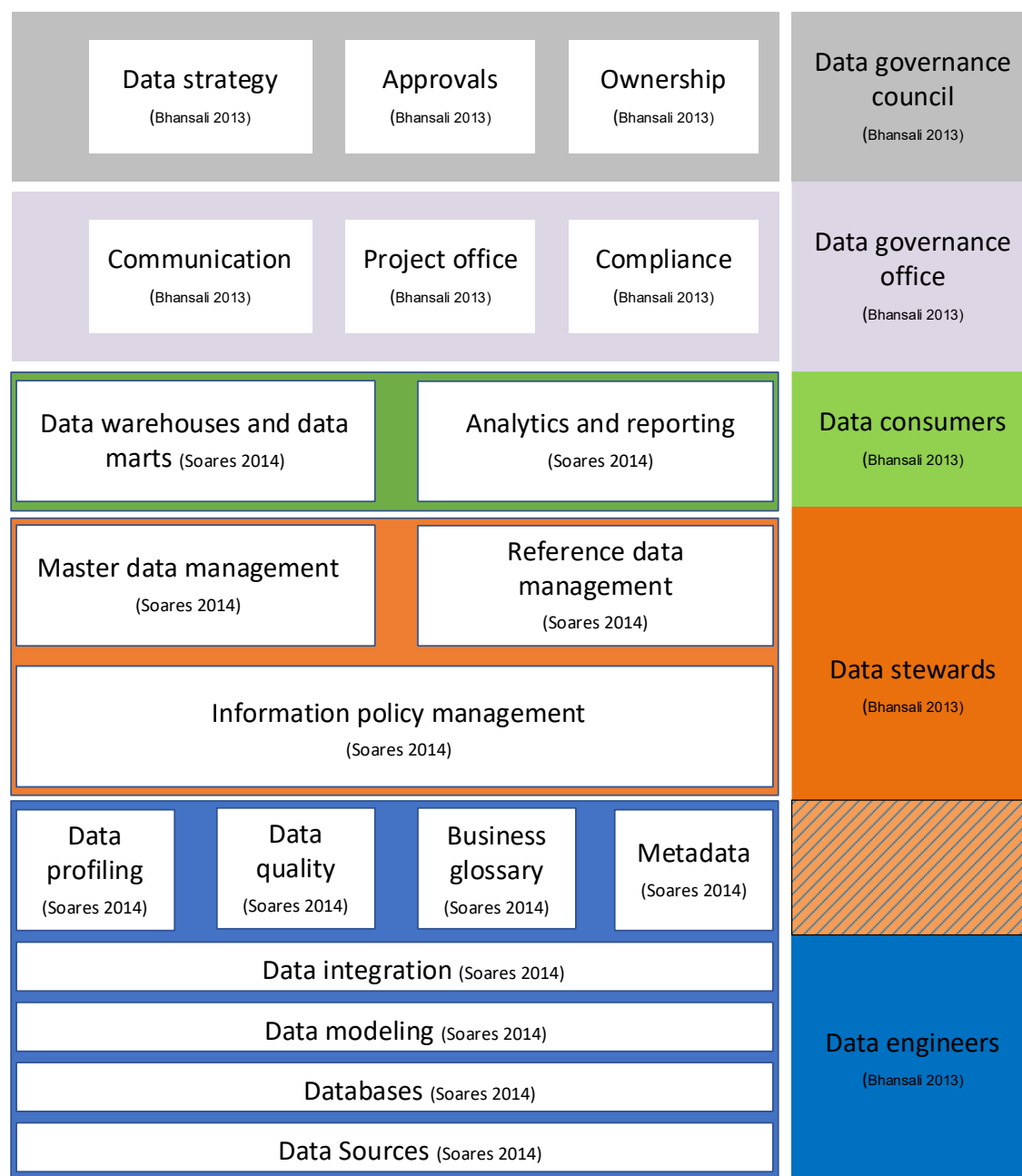


Figure 7-1. The proposed data governance model.

The proposed data governance model points to roles, responsibilities and tasks. Blue is technology stack, orange is data management, green data consumer and top two are compliance, ownership and strategy stack. This proposal was discussed with stakeholders and verified using a set of questions. The questions show that the model is filling the

needed topics for a data governance model. The proposal was finalized to be a better fit for the purpose, with some changes made to the proposed organization and topics for additional study.

## 7.2 Practical Implications and Next Steps

This thesis recommends that a project for planning for a new data governance model implementation should be started. A roadmap for organizational change, and timeline with budget are also needed. Practical implementation will take time. Changing from one operating model to another can be expected to take several years.

Additional steps in developing a modern data governance model for BoF should include:

1. Data strategy is needed
2. Creation of data governance council
3. Creation of data governance office
4. Creation of data steward's role
5. Creation of data engineer's role
6. Data stewards and engineers to identify and create additional processes
7. Data stewards and engineers to agree on data definitions, structure and semantics
8. Data stewards and engineers to agree data profiling, data quality, business glossary and metadata tools and common practices
9. Additional study is needed for data governance technology and architecture stack
10. Additional study is needed for data as a service model, or data factory development
11. Data catalogue is needed, and data portal is to be implemented.

### 7.3 Self-evaluation of the Thesis and Final Words

The business case for the study has become even more relevant since in Finland starting January 2020 when a new information governance law came into effect. The Finnish public sector entities, such as the Bank of Finland, are now obligated to document and report on their information management practices. The objective of the new law is to ensure that the information is managed in a secure way that enables efficient use of the information within the public entity and sharing between public entities, while also allowing general public access to the public information. Thus, sound information governance, and data governance practices are necessary to fulfill the legal requirements, although implementing a new data governance model is most likely going to be difficult.

In this thesis, analysis of the current state of data governance in the Bank of Finland was difficult to conduct. It was evident that there is no common formal data governance, outside compliance and access rights management, in place. Most stakeholders were not willing to openly discuss current challenges. For the relevance of the study, the current state analysis is not that important, however. Making a new proposal was seen as more important. Also, the literature selected for the study helped in building a new data governance model. Literature was mainly and most importantly books and courses and conference materials where used to fill in where necessary.

The study is seen as an introduction to the topic. Data governance is a complicated matter to be covered in single study. Thus, the focus was more on an overview, than giving details. Research design, objective and outcome match as the study is meeting its objective as a proposal, even that additional study is needed on several topics. Also, the stages are not fully linked. More effort was spent on building the proposal, than conducting current state analysis. Finding best practices was crucial to building the proposal and the proposal can be traced back to the literature sources. Finally, validation with stakeholders has been limited. It was difficult to get feedback on the proposal, due to time constraints and lack of subject experts.

As a conclusion, the thesis found that there are no 'correct' answers to building a modern data governance model. Currently, only best practices can be found, implementations are rare. Selecting different source materials should lead into same basic findings with some variation since this study is based on multiple sources. But the basic model, way

of working, roles and responsibilities are well agreed with in the literature, although technology and architecture are open for rapid changes, as they tend to develop at a fast pace.



## References

Bank of Finland [27 April 2020] Bank of Finland. Available from: <https://www.suomenpankki.fi/en/bank-of-finland/> [Accessed 27 April 2020]

Bhansali, Neera (2013), Data Governance, Auerbach Publications

Ferguson, Mike, Enterprise Data Governance and Master Data Management Course, November 2018

Hessing, Ted [30 June 2020] Process Map and Flow Chart Symbols Available from: <https://sixsigmastudyguide.com/process-map-and-flow-chart-symbols/> [Accessed 30 June 2020]

Isbell, Hardin, Underwood (1999), Release 99-113, NASA Available from: <https://mars.nasa.gov/msp98/news/mco990930.html> [Accessed 26 April 2020]

Judah, Saul [19 November 2018] The essentials of data and analytics governance  
Gartner webinar

Keith, Gordon (2007), Principles of Data Management, British Informatics Society Limited

Kharkovyna, Oleksii [7 May 2019], Who Is a Data Engineer & How to Become a Data Engineer? Towards data science Available from: <https://towardsdatascience.com/who-is-a-data-engineer-how-to-become-a-data-engineer-1167ddc12811> [Accessed 23 June 2020]

Ladley, John (2012) Data Governance, how to design, deploy, and sustain an effective data governance program, Elsevier

Limited, A (2019), ITIL Foundation: ITIL 4 Edition, The Stationery Office Ltd, London.

- Linthicum, David [15 November 2016], Cloud computing's next gambit: Data as a service, Infoworld Available from: <https://www.infoworld.com/article/3141461/cloud-computings-next-gambit-data-as-a-service.html> [Accessed 26 July 2020]
- Martinez, Antonio Garcia (2019) No, Data Is Not the New Oil. Wired Available from: <https://www.wired.com/story/no-data-is-not-the-new-oil/> [Accessed 26 April 2019]
- McCombes, Shona [5 June 2020], How to create a research design, Scribbr Available from: <https://www.scribbr.com/research-process/research-design/> [Accessed 9 June 2020]
- Nelson, GS (2018) The Analytics Lifecycle Toolkit: A Practical Guide for an Effective Analytics Capability, John Wiley & Sons, Incorporated, Newark.
- Newman, Daniel [2 March 2017], Data as a Service: The Big Opportunity for Business, convergetechmedia Available from: <https://www.convergetechmedia.com/data-service-big-opportunity-business/> [Accessed 26 July 2020]
- Newman, Daniel [27 June 2017], Why The 'As-A-Service' Model Works So Well For Digital Transformation, Forbes Available from: <https://www.forbes.com/sites/danielnewman/2017/06/27/why-the-as-a-service-model-works-so-well-for-digital-transformation/#ade53ad6490b> [Accessed 26 July 2020]
- Oestreich, Thomas W., Tapadinhas, Joao [20 December 2016], Best Practices for Driving Successful Analytics Governance, Gartner
- Research Approach [9 June 2020] Research Methodology Available from: <https://research-methodology.net/research-methodology/research-approach/> [Accessed 9 June 2020]
- Sarkar, Pushpak (2015), Data as a Service: A Framework for Providing Reusable Enterprise Data Services, Wiley

Sarsfield, S (2009) Data Governance Imperative, IT Governance Ltd, Ely.

Soares, S (2014), Data Governance Tools: Evaluation Criteria, Big Data Governance, and Alignment with Enterprise Data Management, MC Press, Ketchum.

Taylor, Kelsey, Data Governance Maturity Model Available from:  
<https://www.hitechnectar.com/blogs/data-governance-maturity-models-explained/> [Accessed 20 June 2020]

The Economist (2017) 'The world's most valuable resource is no longer oil but data'  
Available from: <https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data> [Accessed 26 April 2020]

White, Andrew (2014) The Gartner Enterprise Information Management Framework, Gartner Available from: [https://blogs.gartner.com/adrew\\_white/files/2016/10/On\\_site\\_poster.pdf](https://blogs.gartner.com/adrew_white/files/2016/10/On_site_poster.pdf) [Accessed 23 June 2020]

White, Andrew, Logan, Debra [28 July 2016], Governing the Information Governance Board, Gartner