

Master data management

A study of challenges and success factors at NCC

Jonas Eyob



KTH Electrical Engineering

Degree project in
ICS
Master thesis
Stockholm, Sweden 2015

XR-EE-ICS 2015:006

Abstract

As enterprises continue to grow, as do the quantity of data and the complexity of storing the data in meaningful ways, for which companies can later capitalize on (e.g. using analytical systems). For enterprises with an international presence it is common that a customer may also be a customer and supplier across borders and business areas. This complexity has created a number of challenges in the ways many enterprises has previously managed their data and is especially important to addresses for some of the enterprises most critical data entities (e.g. customer, supplier, chart of accounts, etc.) also known as *master data*. Master data management (MDM) is an umbrella term for a set of technology independent disciplines focusing on master data with the common goal of helping enterprises to achieve a single version of the truth for its master entities, which in today's enterprises are usually spread across heterogeneous data sources which has historically made it difficult for some enterprise to achieve a collective view on, for example, who their customers are and if they are really meeting the market demands.

The aim of this thesis to understand what challenges enterprises faces when attempting to implement MDM with a primary focus on the governance aspect. An empirical study was conducted in collaboration with NCC AB consisting of interviews with NCC employees, external consultants and literature study. The data obtained from the empirical study where then used as basis for constructing a set of causal graphs to depict the issues identified and the causal factors impacting them; along with recommendations obtained from the external consultants on how to address some of these issues. The purpose is to deduce a number of success factors on how to overcome these challenges in order to NCC to succeed with their MDM initiative.

Results shows that MDM is still a somewhat new concept and where literature on the subject is many times insufficient in details making it hard to understand how to begin. Furthermore, from interviews with NCC employees three dominating problem areas were identified *Common data definition*, *Success change initiative* and *Data quality*, suggesting that most challenges involved in implementing MDM is not technology related but rather organizational and governance related. It is imperative that the business is the one that drives the efforts within MDM initiative as it is the business who best knows the problems they are facing, and how better management of data can improve their ability to realise their business objectives. Furthermore, there needs to be an appointed governance organization which will be responsible for instituting appropriate policies, metrics and roles with clearly defined areas of responsibility in order to instil a sense of accountability for the master entities across business areas.

Sammanfattning

I takt med att företag fortsätter att växa, så följer också volymen av data och komplexitet i att lagra de på meningsfulla sätt för företaget att senare kapitalisera på (t.ex. genom diverse analysverktyg). För företag med en internationell närvaro är det också vanligt att en kund också är en leverantör till företaget, över gränser och affärsområden. Den komplexitet har medfört ett antal utmaningar i sättet företag tidigare hanterat sitt data, speciellt sant är detta för några av företagets mest kritiska data entiteter (t.ex. kund, leverantör, kontoplanen, etc.) också kända som master data. Master data management (eller MDM) är en paraplyterm för mängd av teknologioberoende discipliner med fokus på master data, med det gemensamma målet att hjälpa företag att nå ”en sanning” för dess master data entiteter. Vilket hos företag i dag finns utspridda över heterogena data källor, något som historiskt sett gjort det svårt för de att få en samlad vy av t.ex. vilka deras kunder är och om företaget verkligen möter efterfrågan som finns på marknaden.

Detta arbete syftar till att identifiera några av de brister och utmaningar NCC kan komma behöva adressera för att lyckas med ett potentiellt införande av MDM i organisationen. För att undersöka detta genomfördes en empirisk studie på NCC AB, där intervjuer hölls med både NCC anställda och externa konsulter. Där intervjuer med NCC anställda syftade till att förstå vilka utmaningar och tillkortakommanden som uppkommit i samband tidigare initiativ relaterade till hanterandet av masterdata och nuvarande hantering av master data. Intervjuerna med de externa konsulterna syftade i sin tur till att få en praktisk syn på hur några av de identifierade bristerna och utmaningarna kan/bör bemötas i enlighet med industri praxis.

Med stöd från det empiriska material som samlats in, skapades kausalgrafer för att få en bättre överblick av de problemen och orsaksfaktorernas samverkan.

Resultatet visar att MDM fortfarande är ett relativt nytt begrepp och där litteratur på ämnet i de flesta fall saknar tillräcklig detaljrikhet vilket gör det svårt att förstå hur man påbörjar ett arbete som MDM. Från intervjuer med anställda på NCC har tre dominerade problemområden identifierats *Gemensam data definition*, *Framgångsrik förändrings initiativ* och *Datakvalitet*, något som tyder på att utmaningarna som NCC står inför inte teknikrelaterade utan styrnings och organisations relaterade. Det är således viktigt att affären är de som driver MDM initiativet, då de är de slutliga konsumenterna av data och därför också har bäst insikt i hur förbättrad hantering av master data kan hjälpa de att, på ett bättre sätt, realisera affärsstrategin.

Därför rekommenderas att en styrfunktion inrättas som ansvarar för införande av lämpliga riktlinjer och mätetal såväl som roller med klara ansvarsområden och skyldigheter gentemot de olika master entiteterna.

Acknowledgements

This thesis has been one of the most rewarding experiences during my time at KTH and it is only fitting that I with this thesis at last conclude my studies. Although I will always remain a student, formally, I will no longer have to worry about any exams which looking at it retrospectively I will come to miss.

I want to give my gratitude to my supervisor Mathias Ekstedt from the Royal Institute of Technology for all of your feedback, guidance and kind support during this thesis. It has indeed been a great pleasure for collaborating with you on this.

For my supervisor at NCC, Tomas Elfving, I thank for giving me the opportunity to conduct my thesis for them and for providing me the opportunity to take on, for the organization, a highly interesting and important issue. For all the provided tools and support I am also deeply grateful and for all the people I had the chance to meet and speak with during my time at NCC, I want to send my deepest thanks.

Last but certainly not least I want to thank my family for always believing in me and supporting me through ups and downs - Thank you!

List of abbreviations

MDM	Master data management
PMO	Program management office
CIO	Chief executive officer
CMO	Chief marketing officer
DG	Data governance
DQ	Data quality
MDGC	Master data governance council
EA	Enterprise architecture

List of figures

Figure 1 NCC merger and acquisition history	2
Figure 2: The NCC organization.....	5
Figure 3: Group IT organizational chart	6
Figure 4 The Governance V (Ladley, 2012)	11
Figure 5: Data governance model adopted from (Khatari & Brown, 2010)	13
Figure 6 Common Pitfalls of MDM adopted from (Radcliffe, 2011)	18
Figure 7 Common Pitfalls of MDM adopted from (Radcliffe, 2011)	18
Figure 8. Induction and deduction, adapted from (Wiedersheim-Paul & Eriksson (1991) cited in (Le duc, 2007)	22
Figure 9 Structure in which data was entered during coding	26
Figure 10 Example of a consolidation during coding	26
Figure 11: Identified problem and challenge areas	29
Figure 12: Opinion graph - Common data definition.....	30
Figure 13: Opinion graph - Successful change initiative	31
Figure 14: Opinion graph - Data quality	34
Figure 15. Common data definition - proposed solutions.....	36
Figure 16: Governing Structure as derived from interview with Consultant 3	37
Figure 17: Opinion graph (solutions) - Successful initiative	37
Figure 18: Opinion graph (solution) - Data quality.....	40
Figure 19: An example of deriving metrics from strategic objective.....	47
Figure 20 Governance structure derived from (Cervo & Allen, 2011; Loshin, 2009; Dreibelbis et al, 2008)	52

Table of contents

Abstract.....	2
Acknowledgements.....	4
1 Introduction.....	1
1.1 Problem definition.....	1
1.2 Thesis objective.....	3
1.3 Project contextualization.....	3
1.4 Delimitations.....	3
2 NCC AB.....	5
2.1 NCC Strategy.....	6
2.2 Group IT.....	6
3 Literature study.....	7
3.1 Master data.....	7
3.2 What to master.....	8
3.3 Master data management.....	9
3.4 Master Data Governance.....	10
3.5 Change Management and MDM.....	17
3.6 Data Quality Management.....	19
3.7 Data stewardship.....	19
4 Methodology.....	21
4.1 Research Approach.....	21
4.2 Research processes.....	22
4.3 Interviews with external expertise.....	27
4.4 Method validity and generalization.....	28
5 Empirical findings.....	29
5.1 Interviews.....	29
5.2 Observation.....	40
6 Discussion.....	42
6.1 “What shortcomings and challenges are there in the way master data is managed at NCC today and what causes them?”.....	42
6.2 “With regards to the identified shortcomings and challenges which are the success factors enabling NCC to improve their management of master data?”.....	44
7 Conclusion.....	53
7.1 Recommendations.....	53
7.2 Reflection.....	54
7.3 Future work.....	55
8 References.....	56
Appendix.....	59
A.1 Role description.....	59
A.2 Deloitte’s master data identification framework.....	60

1 Introduction

In today's rapidly changing business environment, the emphasis on enterprise agility is becoming increasingly important. Where business executives, are looking for quicker and improved decision making by harnessing the data stored in the enterprise by means of data warehouses and business intelligence tools. However, for most seasoned enterprises with business areas having evolved with business applications focused on vertical success (Loshin, 2009), which with time have given rise to "information silos" often with redundant data stored in overlapping systems. However, while this may not introduce any concern seen from the perspective of each business, this however give rise to a range of challenges when consolidating data on enterprise level. Years of operating in silos have consequently resulted in a range of differing definitions across business areas of otherwise identical data causing data quality problems common in many enterprises.

Master data management (MDM) is a branch within the field of Enterprise Information management (EIM) focusing on master data (Berson & Dubov, 2011). It is comprised by a combination of processes, standards, governance, organization and technologies with the objective of providing consumers of master data across the enterprise with a single – authoritative - view of its master data. Many times an organizations master data is scattered across a number of disparate data sources. Where master data consist of those key information objects for which maintaining good quality is critical as they form the basis for business processes (Loser, 2004).

Put short the idea of MDM is ultimately to centralize an enterprises master data and make it accessible through a unified view, with agreed on data definitions.

However in order for enterprises to be able to realise the benefits of MDM it is imperative that governing processes are in place to control the use, resolve conflicting interests and continuously measure data against a set of agreed on quality dimensions. For this reason an integral part of MDM will be concerned with matters such as data governance (DG) and data quality (DQ).

Although literature on the subject is rather scarce, it has been established MDM is not solely a technical implementations, quite the opposite, in fact many of the toughest and most challenging aspects of MDM lies with organizational and governance issues, which is further strengthen by Radcliffe (2007) who puts forwards seven building blocks essential to MDM, one of them being technology. For many enterprises exploring the opportunities and benefits that can be obtained from MDM are done so with little regards to the "soft parts".

Through improved managing of the enterprises master data enterprises can make better informed decisions, understand its customer better by having a collective view of its customer across the enterprise for the benefits of coming "closer" to the customer, to name a few. Furthermore, MDM also helps to dissolve many of the inherent complexities in data synchronisations, and through an increased standardization in definitions there are increased opportunities for operational efficiencies, as well as consolidation and elimination of redundant data.

1.1 Problem definition

NCC is one of the leading construction companies in the Nordic region with more than 18 000 employees and an estimated turnover of 56 BSEK as of fiscal year 2014. NCC's offerings ranges from property development for both commercial and private use, road services, producer of asphalt gravel products and other infrastructure related services. These services are in turn provided by one of the four business areas within NCC, namely, Construction, Housing, Property Development and Roads. NCC first came to existence through the merge of the two construction companies ABV and Johnson Construction Company (JCC) in 1988. Since then NCC has acquired a number of companies as can be seen in Figure 1, which seen from an IT and data perspective have resulted in increasing complexity in the IT landscape as well as

information silos making it hard to achieve a single view of the organizations customers, ongoing projects to name a few.

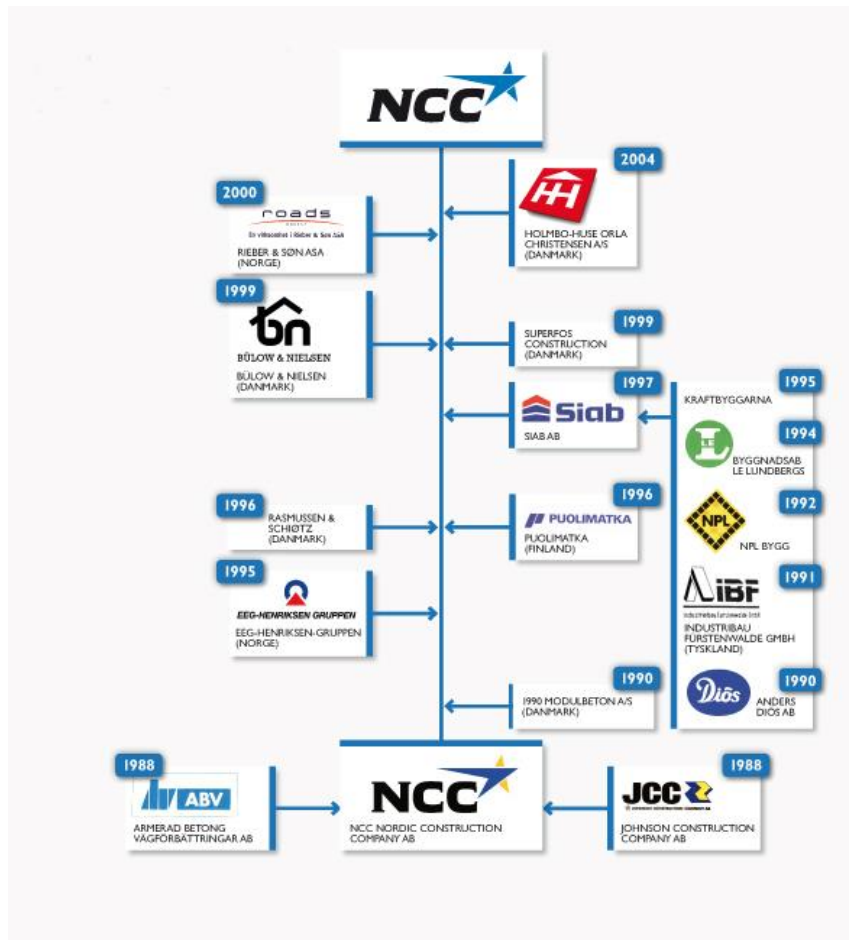


Figure 1 NCC merger and acquisition history

At NCC they have set of common data concepts that are regarded as fundamental and shared by each of NCC four business areas *construction, roads, housing and property development* which are also referred to as their master data objects which include customer, supplier, project, organization and employee. NCC, as with many other enterprises, are struggling with an immature IT and data management practice where the complex IT landscape has introduced some notable difficulties for being able to manage its data. Nonetheless, NCC has started to understand the benefits of IT and its data, and if managed in a structured way can turn out to be a strategically valuable asset.

In an effort to try become more cost-effective as well as try come closer to the customer (Annual report, 2013) the strategic direction “One NCC” has meant that the NCC Group is looking to identify synergies across business areas by consolidating redundant processes. Which in turn have proved to be a challenging task due to the number of acquisitions that have been made and the lack of a unified vocabulary for some of the common concepts, such as: customer, supplier and project.

The importance of IT in its role to support the business has been recognized as evident from the work with enterprise architecture and is still an ongoing program. In this light, the need for improved reporting has been expressed by management, as there no easy way to achieve a collective view of how the various business areas are performing.

One of the critical systems present in the NCC IT landscape today is the master data repository system called Gemensamma Registret (or GR), which houses many of the common data

concepts. However due to historical reasons and previous failed attempts to try address the deficiencies of multiple definitions of the common data concepts (master data), this thesis has as primary objective investigate why previous initiatives with regards to master data initiatives has failed to succeed and against this insight try to propose a set of success factors that would enable them to proceed with its MDM initiative.

Through interview with key employees at NCC the objective is to try gather opinions and testimonials of what problems related to master data that exists and why previous attempts has been unsuccessful. Through the extraction of these opinions a causal opinion graph will be designed from which conclusions will later be drawn.

Research Questions

In order to provide NCC with recommendations on how to proceed with its MDM initiative, from the problem definition, two research questions regarded as critical to gain the needed context of the problem were derived:

- “What shortcomings and challenges are there in the way master data is managed at NCC today and what causes them?”
- “With regards to the identified shortcomings and challenges which are the success factors enabling NCC to improve their management of master data?”

1.2 Thesis objective

- To lift the awareness among NCC employees to the problems around master data and why previous initiatives addressing these problem has failed.
- Propose a reference model for a governance organization describing the needed roles and their areas of responsibility.

1.3 Project contextualization

In (Ross, et al., 2006) the authors states that an effective foundation for execution depends on a tight alignment of between businesses objectives and IT capabilities. For many enterprises it is however the case that new strategic initiatives rarely can be realised without each time needing to implement new IT solutions (Ross, et al., 2006). In some sense this captures the current situation residing at NCC.

In a new strategic effort to get a better overview of the businesses performances for better decision-making NCC are now looking at the possibilities of introducing a new solution for better informative business intelligence tool at group level. A precondition for enabling an increased degree of process standardization and service reuse is having a clean and consistent master data on which the services are later built upon. However, as there is a lack of knowledge as to how to work with master data, this has had consequences not only on data quality, but in the form of increased costs for new systems and lengthy implementations.

The results from this thesis will as seen aid group IT in understanding which challenges lay ahead with respect to a MDM governance function and how this can help them advance towards a “target architecture”; and a way of working so that, rather than being reactive and building IT solutions whenever a strategic initiative changes, building IT capabilities to be proactive. (Ross, et al., 2006).

1.4 Delimitations

Given the limited time frame allotted for this thesis (20 week) some delimitations have been made.

- i) As the concept of MDM covers both business and technology, this thesis less focus is placed on the technology side which covers various implementation styles and their challenges and suitability. Reason for this stems from literature for which

there have been countless references to MDM challenges as laying with organizational aspects.

- ii) While this scope has been narrowed down to address organizational aspects of MDM, there must still be further delimitations to the scope. Based on a few exploratory interviews conducted at NCC (see methodology) and what literature has been emphasized as most crucial when establishing MDM –for this thesis the focus will be on the data governance dimension of MDM.
- iii) Due to limited time-frames and interviews are only performed with employees residing in Sweden, hence, no interviews are performed with representatives from Denmark, Norway and Finland.

2 NCC AB

NCC AB is one of the leading construction and property development companies in the Nordic region with over 18 000 employees and a yearly turnover estimated at SEK 57 Billion. NCC business operations consists primarily three areas that is 1) *Construction and Civil engineering* – this operating sector manages all of NCC construction operations and is divided up into four business areas *NCC Construction Sweden*, *NCC Construction Norway*, *NCC Construction Finland* and *NCC Construction Denmark*. Each of these business areas are privately held subsidiaries companies of NCC AB each with their own organizational structure and CEO. Of these four, NCC Construction Sweden is the biggest of them with respect to employees and turnover which makes up, approximately, 54% of the Group’s overall annual turnover. 2) *Development* – this operating section is managed by the two business areas NCC Housing AB and NCC Property Development AB, where the former developing and selling permanent housing in selected Nordic markets. The latter is responsible for developing and selling commercial properties in growth markets in the Nordic and Baltic region. 3) *Industrial* – this sector is operated by NCC Roads AB whose core business service consists of aggregates and asphalt production as well as asphalt paving and road service.

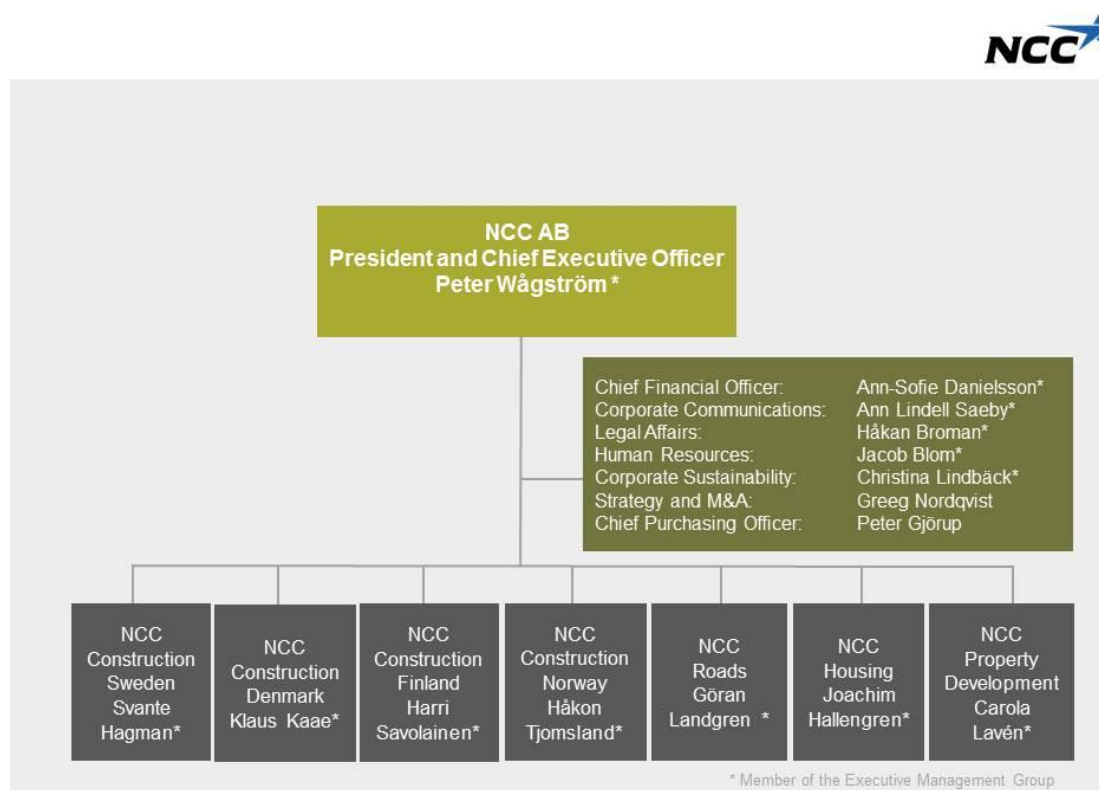


Figure 2: The NCC organization

As seen in Figure 2, the different NCC construction units unlike business units such as NCC Roads, Housing and Property development that have a steering group function for the countries in which they are operating in. This is not the case for the construction units who for reasons beyond the scope of this thesis have no steering top that is common for all the operating countries.

2.1 NCC Strategy

As NCC operates in mature markets, characterized by its price based competition, being able to grow profitably is essential. One ambition of NCC is to become the customers' first choice which is accomplished by maintaining as well as building new customer relations by delivering the right product, with the right quality, on time and for a competing price. To achieve this, the ability to identify cost reduction are critical and furthermore emphasizes the importance of being able to realize synergies across the business areas and follow up on projects to understand where improvements can be made.

On the basis of this, the strategic direction named *One NCC* was formed. In essence this group-wide strategy has articulated the wish to better understand its customers, seize new business opportunities and for the group to come closer and act as "one". This of course requires sufficient system support in order to support the business with the needed data. With companies transitioning towards an increased digitalization of core processes, the requirements set on the data management practice and intelligence systems will become even more critical for reliable data driven decision making.

2.2 Group IT

Group IT is a fairly new function at group level with the purpose of coordinating group-wide IT activities in order to better utilize synergies and maximize cost-effectiveness by having a function taking a holistic view on the IT landscape. Prior to the establishment of group IT each business areas operated within themselves creating information and system silos, which resulted in multiple processes and systems with overlapping information. Today as Group IT is still a function at its early phases, although much progress has been done and continuous work trying to identify common denominators among the business areas that would benefit from being managed at group level is ongoing.

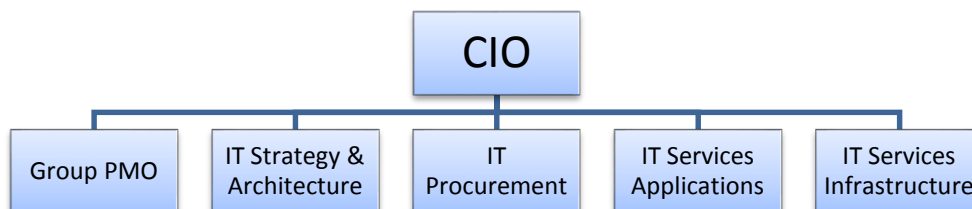


Figure 3: Group IT organizational chart

Group PMO

The Group PMO (Project Management Office) function is responsible for ensuring that project manager and project owner are provided with the necessary tools and framework that will enable them to manage their projects in a cost-effective and value-driven manner.

“To succeed we need to constantly improve our business and be prepared to change our ways of working. To manage changes is a key to success”

- Ann-Sofie Danielsson (CFO), NCC

Besides providing a reference framework for enabling project manager to better understand each other, providing tools and proven methodologies for managing change is also within the scope of PMO. As seen from the quote above, change is critical, especially with regards to the internal projects which has not been as successful as the external (construction) projects.

IT Strategy & Architecture

IT Strategy & Architecture (or Group EA) is responsible for Groups ITs strategic directions, IT architecture and IT security. Through a holistic perspective, Group EA ensures alignment

between business and IT strategy throughout the organization. Today Group EA offers EA services to various stakeholders from business, IT-management and projects in conformance with the “One NCC Project Steering Model”. One of many initiatives taken by Group EA was the establishment of an EA council consisting of IT architects representatives from the various business areas in order to ensure that a holistic view is taken on every IT project.

At present state the EA services consists of following:

- Project Architecture – When embarking on a new IT project the solution architect presents the architecture to the EA council for revision with respect to reusability, alignment to IT strategy and validation against IT policy.
- Solution Selection – To assign solution architects to projects in order to produce a solution that is reusable, meets business requirements and aligns to the overall IT strategy.
- Reuse optimization –Identifying consolidation opportunities and making appropriate business cases for them. Moreover, to avoid further complexity to the IT landscape and excessive costs by preventing adding new IT components that already exist.
- Master data management – To realize a common definition of master data entities, and ensuring master information and data ownership.
- IT Policy and guidelines management – Continuously working to ensure that appropriate IT-policies are managed and complied to.

IT Procurement

The IT procurement function is responsible for the groups IT agreements, but also assists the local business areas with professional procurement services.

IT Services Applications & Infrastructure

The IT Service function is divided up into two camps: *Application* and *Infrastructure*. The application camps are responsible for the development, operating and management of group-wide systems; infrastructure is responsible for the development, operation and management of group-wide IT infrastructure services. They are also in charge for group-wide outsourcing agreements and licenses. Both camps has as primary goals to reduce any overhead costs by reducing sub-optimal solutions and to meet business requirements as agreed on in the SLAs’.

3 Literature study

3.1 Master data

Today, many definitions of what characterizes master data (MD) exist; common for them is the notion of master data as those core business entities used by different applications across the company, along with their associated metadata, attributes, definitions, roles, connections, and taxonomies (Loshin, 2009). In short, they are the enterprises most important business entities which define the enterprise (Dreibelbis, et al. 2008) and are the focus of multiple business processes (White *et al*, 2006; Otto and Hüner, 2009). Some common master data objects (or core information objects) include, but are not limited to: customers, employees, suppliers, products, policies (Dreibelbis *et al* 2008; Loshin 2009; White *et al*, 2006). However not all data is considered master data, and some attempts trying to distinguish master data from other types of company data have been made. In (Wolter and Haselden, 2006) and (Oracle, 2009) six types of data exist in companies:

- **Unstructured.** This is data found in email, white papers like, magazine articles, corporate intranets portals, product specifications, marketing collateral, and PDF files.
- **Transactional.** This is data related to sales, deliveries, invoices, trouble tickets, claims, and other monetary and non-monetary interactions.

-
- **Metadata.** This is data about other data and may reside in a formal repository or in various other forms such as XML documents, report definitions, column descriptions in a database, log files, connections, and configuration files.
 - **Hierarchical.** Hierarchical data stores the relationships between other data. It may be stored as part of an accounting system or separately as descriptions of real-world relationships, such as company organizational structures or product lines. Hierarchical data is sometimes considered a super MDM domain, because it is critical to understanding and sometimes discovering the relationships between master data.
 - **Analytical.** This is data used for enterprise performance calculations and to aid in companies' decision making. More often than not, this type of data is stored in Data Warehouses with various capabilities for further aggregation and analysis (Oracle, 2009).
 - **Master.** Although master data is non-transactional they are used within transactions, across different applications. Master data can be further grouped into four categories: people (customer, employee, sales person), things (divisions), places (office locations, geographic divisions), and concepts (contract, warrantee, and licenses).

Otto and Huner (2009), condensed into four dimensions, further described how master data differs from other types of data:

- Master data, unlike inventory and transactional data, captures characteristics from real world objects.
- Parts of the master data object will through its lifetime remain static. That said, additional attributes could be added in case it is needed, but this will not affect existing base data.
- Instances of a master data record (e.g. data on a particular customer) will, unlike transactional data, remain constant with respect to volume.
- Master data can exist by itself without the presence of transactional data and constitutes the reference to transaction data, the opposite do however not hold.

Essentially, the point of using master data is to bring the organization closer by sharing these data objects across business divisions. Furthermore, with shared master data comes also process simplification and uniformity of processes (White, et al., 2006).

3.2 What to master

One of the challenges involving master data and the management of it is to understand which data should be mastered. There are different techniques can be employed when deciding what data to master, Wolter and Haselden (2006) suggest a framework entailing eight criteria for deciding on whether data should be mastered. Similar to the aforementioned approach is the structural scoring framework proposed by Deloitte (see Appendix A.2) where a data is evaluated against three criteria: *shared* – is it used by more than one business process/system; *value* – how fundamental is the data for the business; *volatility* – what is the data modification behavior of the data. Based on the received score the data is then either qualified as master data or requires further study.

The use of frameworks for identifying master data does however not provide a single truth in terms of what later on is actually mastered or not. Wolter and Haselden (2006) explains that companies may still need to master some data that may not qualify as master, and likewise, data qualifying as master data may not be managed as such. Rather than simply enumerating entity types as candidates for being mastered, these frameworks should be used within the context of a business need to gain a better understanding for data behavior and criticalness (Wolter & Haselden, 2006).

3.3 Master data management

3.3.1 Master data management (MDM): Definition

The management of master data is not new, in many organizations systems for storing and retrieving business critical data exists (Dreibelbis, et al., 2008). Since these master data systems originally were constructed in support for a specific line of business, when enterprises grew so did the complexity of the IT-landscape. These, otherwise sufficient, homegrown systems consequently struggled to provide a consistent view of the enterprise's master data objects. As enterprises IT landscape gradually became more complex as did any attempts of master data integration, hence, it may have been considered easier to create new applications and databases to facilitate changing strategic initiatives, rather than modifying existing applications (Dreibelbis, et al., 2008).

According to David Loshin (2009) master data management (MDM) comprises a set of data management best practices to help corporations' key stakeholders and business clients with incorporating business applications, information management methods, and data management tools in an endeavour to improve and maintain clean, accurate, timely and complete data across disparate applications (Butler, 2011; Kernochan, 2006). According to (White, 2007; Dreibelbis, et al., 2008) one of the main objectives of MDM is to provide authoritative master data to an enterprise known as a *system of record* (SOR). The SOR is the one source where master data is guaranteed to be accurate, up-to-date and thus the best source of truth (Dreibelbis, et al., 2008; White, 2007).

MDM systems can, depending on the consumer, be categorised in to one of three key patterns of use (Dreibelbis, et al., 2008):

- **Collaborative Authoring.** As authoring of a master data object may be conducted by several people, this gives rise to a highly complex workflow, in where a multitude of topics must be agreed on by a group of people (Dreibelbis, et al., 2008). Typical example is that of product information management (PIM) systems. Due to the complexity in developing and managing products, requiring several parties to agree, PIM systems commonly support the collaborative style (Dreibelbis, et al., 2008). Furthermore, for a collaborative MDM style to execute effectively this requires core capabilities such as task management, and state management to guide and monitor tasks being collaborated on. Since task may be worked on concurrently there is a need for control mechanisms to preserve data integrity (Dreibelbis, et al., 2008).
- **Operational MDM.** According to (Butler, 2011) is a solution intended to manage transactional data used by operational applications. This is realised by providing stateless services that can be invoked by business processes or directly by applications. The nature of the operational MDM thus makes it suitable to be incorporated into a service-oriented architecture environment (Dreibelbis, et al., 2008).
- **Analytical MDM.** As sound and strategic good decisions are of paramount importance for any organization, analytical MDM has surfaced as means of managing the enterprise's analytical data (see §2.1 Master data). The goal of analytical MDM is, as cited by (Butler, 2011) "...providing high quality dimensions with their multiple simultaneous hierarchies to data warehousing and BI technologies". As further explained by (Dreibelbis, et al., 2008) analytical MDM is essentially about the intersection with Business Intelligence (BI) of which three such intersections are (i) trusted data source (ii) analytics on MDM data and (iii) analytics as a key function in a MDM system. For BI tools to be able to provide meaningful analysis the data on which these tools operate must be good and trustworthy. Furthermore, as analytical MDM systems provide to means of improving the data quality in a heterogeneous application landscape these shortcomings in data quality will also make its way in to the BI tools.

3.4 Master Data Governance

3.4.1 Governance: definition and motivation

According to Dreibelbis et al (2008) governance is the decision and action processes of managing shared resource for common good, which includes:

- Identifying shared resources that would benefit from being managed.
- Communicate who is responsible for a resource; how the governing bodies are organized and the processes they will follow.
- Distributing decision making with regards to what is encouraged and likewise discouraged when touching upon shared resources
- Assigning roles and responsibilities to different parts of an implementation involving processes and policies for a shared resource.

Corporate governance, as defined by the Organization for Economic Cooperation and Development (OECD), also provides the structure through which company objectives are set and their performance monitored, as to make sure they are attained (OECD, 2004). Governance is in nature a highly complex process which may be executed depending on context; this is further underpinned by OECD (2004) who emphasizes that there is no single good corporate governance model.

Weill and Ross (2004) recognized that in order for effective governance to be achieved clear desirable behaviors, embodied in the organizations beliefs and principles needed to be set. Further they also observed that enterprises with a common mechanism to handle the key assets (e.g. human assets, financial assets, physical assets, IP assets, information & IT assets and Relationships assets) performed better than those who without proper governance for mentioned assets. Moreover, they argue that if the same executive committee governs both financial and IT assets, this would likely result in benefits such as increased integration and subsequently create more value by leveraging on synergies between these.

There is consensus within the MDM community with regards to the importance of data governance (Loshin, 2009; Berson and Dubov, 2011; Tuck, 2008; Ballard, et al, 2013), with Radcliffe (2007) explaining that without effective data governance MDM initiatives are deemed to fail. This is further strengthened by a study conducted by PwC (Messerschmidt & Stüben, 2011) which showed that only 27 % considered implementing a state-of-the-art MDM solution to be key for MDM, instead governance and good management accounted for 71% and 69% respectively.

3.4.2 Data Governance Fundamentals

A continuation of the otherwise broad concept governance – focusing on data – is that of data governance (DG). Many attempts trying to capture the essence of data governance in few sentences has been made and listed below are excerpts of them:

- *“..is a set of processes that ensures that important data assets are formally managed throughout the enterprise”* (Sarsfield, 2009)
- *“..is expected to ensure that the data meets the expectations of all the business purposes in the context of data stewardship, ownership, compliance, privacy, security, data risks, data sensitivity, metadata management, and MDM.”* (Loshin, 2009)
- *“..is a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods”* (Data Governance Institute, u.d.)
- *“Data Governance (DG) – The formal orchestration of people, processes, and technology to enable an organization to leverage data as an enterprise asset.”* (The MDM Institute, u.d.)

As seen, a plethora of definitions defining the concept of data governance exists. Bottom line being that data governance has less to do with the stored data itself but rather with the roles, policies and processes; guiding and enabling ownership and accountability for data (Plotkin, 2014). Data governance help to organize roles properly and advocates a way of working that will make data understandable, trusted, and of high-quality (Plotkin, 2014). Through proper data governance ownership of data is instituted which in turn will instill a sense of accountability and responsibility for ensuring data is of good quality and that guidelines and policies for data are enforced.

One of the most common definitional mistakes coming to data governance is the perception that data governance (DG) and data management (DM) is the same thing, albeit expressed differently. This is however not the case according to (Dyché and Nevala, 2009; Ladley, 2012); who argues that DG and DM are two sides of the same coin as illustrated in Figure 4.

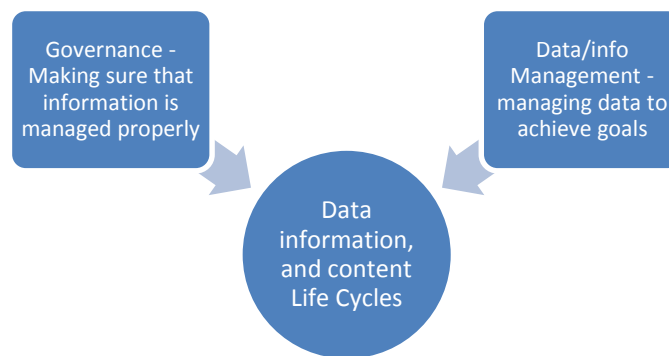


Figure 4 The Governance V (Ladley, 2012)

From Figure 4 it is understood that data managers are those responsible for adopting and executing the policies enforced provided by the DG function. The DG function can thus be regarded as that of an auditor, which is, reviewing the work of data managers to ensure that management of data is conducted according to policies. Furthermore, Dyche & Nevala argues that since DM is an IT-function they should ideally report to the CIO, unlike the DG function which is per definition a business driven function and should therefore report accordingly.

Data Governance scope

In any large organization there will be an overwhelming amount of data, each having their part in the day-to-day business tasks. However, there will always be some data what are more critical to the organization, and that will thus require special attention –that is they need to be governed. When deciding on the scope of the DG, it is as reported by Ladley (2012) not practical to govern all type of data. Instead there are three parameters to consider when deciding on scope (Ladley, 2012):

- (i) Business model
- (ii) Content
- (iii) Degree of federation

In any organization there may be several different lines of businesses (LOB) each with data that are dear to them. There can be data that are shared across many or all LOBs which would then imply that the scope of the DG would have to encompass all LOBs; if however data is completely different then this might require separate DGs for each LOB (Ladley, 2012). With regards to content parameter this refers the question of “what content in the organization is to be governed?”. Typical data types subjected to data governance are BI data, master data and other structured data.

Data governance foundation

Data governance as a practice is highly mutable, that is, the structure of the DG function to large parts depends on the company it is being deployed at. Weber et al (2009) furthermore argues that previous data governance structures has neglected the fact that companies require specific configuration to best align with a set of contingencies; and therefore, it is imperative that a company's DG structure be designed taking this into account; as contingency theory argues that an organization's traits and its effectiveness are determined by its contingencies (Weber, et al., 2009).

Although there are certain core areas of DG program that nevertheless must be decided upon. Ladly (2012) proposes six areas:

- Organization –formulated roles for enabling accountability and responsibility. Issue resolution is critical as data governance is an intersectional practice therefore having the necessary decision structure can be critical.
- Principles – an organizations core beliefs (high-level opinions about data) out of which all other decision are based. In several studies conducted by Weill and Ross (2004), they had identified that enterprises with superior results where those with a set of stated IT principles. Although this was recognized within the context of IT governance, Khatri and Brown (2010) argued that wisdom acquired from IT governance can be transferred onto Data governance.
- Policies – enforceable processes that is a codification of a principle. Unlike policies, principles are considered too abstract to enforce and through policies the principles are made more concrete.
- Functions – describes which series of action must be performed early in the DG program. Ladley argues that the role of function is twofold 1) to bring an awareness of what must be done 2) since the needed actions (functions) has been identified this in turn aids in assigning groups or individuals accountable and responsible for critical areas.
- Metrics – being able to display improvement of data is crucial, especially in a organization where top management may not see the immediate value in the DG program. Thus, having some carefully selected metrics are vital for a sustainable DG.
- Technology – having systems to support the daily tasks of DG can be helpful. Although technology must not be purchased before a DG program has been deployed as technology is mere but a means to effectivize areas that has been identified to benefit from autonomy.

Despite the lack of consensus in many areas of DG, Otto (2011) reportedly identified three points that, among practitioners, data governance must find answers to:

- *"What decisions, with regard to corporate data, need to be made on an enterprise wide level?"*
- *"Which roles are involved in the decision-making processes"*
- *"How are the roles involved in the decision-making process"*

Khatri and Brown (2010), drawing inspiration from the work of Weill and Ross (2004), lists five "decision domains" depicted in Figure 5, to answer the first question posted by Otto (2011).

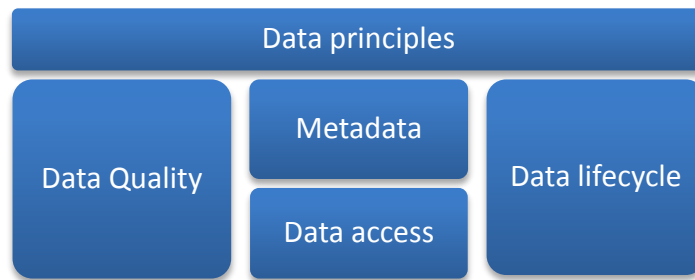


Figure 5: Data governance model adopted from (Khatari & Brown, 2010)

As depicted in Figure 5 - *Data principles* – is placed at the top, this is since data principles is the linkage to business, and wherein role of data as an asset must be communicated out to the business leaders. The field of data quality and its impact on enterprises effectiveness is topic that has been visited countless times (Redman, 1998; Loshin, 2009). According to Khatari and Brown the costs of poor data quality for USA business costs 611 Billion every year in postage, printing and staff overhead; and a problem still not given the sufficient level of attention (Haug, et al., 2011) – data quality will be discussed more thoroughly in the following sections. Metadata is the data about data, there to help interpret the data. Metadata can be differentiated in to different domains depending on the data it describes, namely: physical (e.g. data describing the type of attributes in a database table), domain-independent (that could be descriptions of creator & modifier of data or access rights information to data), domain-specific (describes which division specific data resides in the organization), user metadata (e.g. this could be data commonly expected or associated with a data item such as usage history), data access (integrity, confidentiality and availability of business data is assessed by security officers who suggests safe guards to maintain these. On the basis of this data access policies and alike are formulated) and lastly data lifecycle (understanding a data’s usage pattern and different stages in its lifecycle is important as this enables the organization to optimize the way data is stored, which could have a positive impact on costs).

With regards to the second point, Cervo and Allen (2011) emphasize that roles be clearly defined in the DG model. Furthermore, while deciding on who to appoint a certain role one should not assume that roles and title be used interchangeably. Although it is necessary to involve individuals that are able to influence decision and that have decision mandates to push decisions through, it is equally important that the appointed individuals to these roles are “doers”. And since the structure of the DG organization to a large degree depends on the particular company there is generally no generic model that is guaranteed to work. However, there are some key roles that continuously tends to reappear in literature and who (Sarsfield, 2009) calls key roles to companies looking to jumpstart their data governance program.

Executive sponsor – the role of the executive sponsor is to make sure that DG initiative has board level backing (Loshin, 2009; Sarsfield, 2009), which implies that the executive sponsor commonly holds a C-level position within the company. Who the sponsor(s) should be, as Sarsfield (2009) explains, depends on the data that is causing the most problem for the business. For example, problems with compliance issues and perhaps master entities such as charts of accounts, is likely of great interest for CFO; while, for example, a Chief Market Officer (CMO) would likely be the sponsor in case problem lies with customer data.

Data Governance council (Strategic) – consists of directors or senior manager level representatives from different LOBs (Loshin, 2009; Patton and Wentz, 2010; Cervo and Allen, 2011) and are responsible for defining policies, standards, quality agreement, priorities and communicating out to the enterprise their role and how this impacts the day-to-day processes. Moreover, Cervo and Allen (2011) stresses the importance of selecting council members with sufficient influence and authority over functional teams, who eventually are creators of master data.

Loshin (2009) suggest that this council be chaired by the *data governance director* role, who is essentially in charge for the day-to-day management of enterprise data governance. Moreover

he is also responsible for overseeing conformance with information policies and regulations, and to provide periodic reports on the data governance performance.

Data Stewardship council (Tactical) – operates under the direction of the data governance council and consists of interested stakeholders from across the enterprise (Loshin, 2009). As seen such as business managers. This council can be seen as the governance council’s “extended arm”, responsible for implementing policies put forth by the governance council, by developing procedures and standards for being able to meet data quality requirement. It is then the responsibility of each data steward accountable for either all or parts of a data entity to make sure it is correctly managed.

The data stewardship council is also responsible for overseeing the work of data stewards, to ensure data quality requirements and policies are continuously being met (Loshin, 2009).

Data Stewards – are responsible for everyday operations and to enforce policies and guidelines from the data stewardship council (Ballard, et al., 2013). Sarsfield (2009) calls them the “technologists” referring to their roles as providing support with systems and data access, as well as some more technical task such a making metadata mappings. Loshin (2009) on the other hand, argues that data stewardship is not necessarily an IT function as oppose to Friedman (2007) who explicitly states that stewards should reside in the business and not in the IT organization.

Data stewards are typically assigned by the stewardship council based on either subject area or LOB (Loshin, 2009). However, within the context of MDM master entities may span across multiple LOB, and so, rather than having a single data steward accountable for a master data entity, the stewardship role is aligned along master data boundaries.

3.4.3 The role of Data Governance in MDM

To understand the role of data governance one must understand the challenges associated with implementing an MDM program. As an MDM initiative will incrementally impact the whole enterprise this is an enterprise-wide concern involving business managers from different business areas. This intersection of people, business processes, information and systems is what makes MDM such a complex matter (Cervo & Allen, 2011). In turn this then calls for a structured methodology for addressing issues such as ownership, accountability, policies and appropriate roles to overcome business local self-interest to benefit the overall enterprise - which is where data governance emerges (IBM, 2007). Data governance is imperative for successfully launching an MDM program, which is also seen from a study conducted by “The Information Difference” (The Information Difference, 2010) involving 257 companies participating in the study to understand how data governance links to MDM and data quality, where 48 % percent of the companies surveyed considered it a good idea to implement DG before attempting to implement MDM. Furthermore, Cervo & Allen (2011) argue that initiating DG prior to MDM is fine, the opposite however is not; there can be no effective MDM practice without DG – DG in the context of MDM can be regarded as the “glue” keeping everything together (Cervo & Allen, 2011).

Since one of the drivers for embarking on a MDM program is to improve the overall data quality, DG will be of paramount importance to realize this. From the study conducted by the Information Indifference, results found that 93% of those planning to implement data governance did this with the intention of measuring data quality.

3.4.4 (Master) Data Governance challenges and considerations

The importance of DG within MDM is evident seen from literature, where in essence all of them emphasis that success of MDM depends on DG, and without one can neither implement

nor operate MDM (Loshin, 2009; Berson and Dubov, 2011). DG is considered the “glue” within MDM keeping all the needed pieces together (Cervo and Allen, 2011).

Despite its arguably important role, DG is not easily adopted by enterprises and there are a myriad of challenges and factors that must be considered in order to succeed with DG.

- *A business function.* Data governance, unlike IT governance, is the involvement of the business (Dreibelbis, et al, 2008). Since it is the business that eventually consumes and owns the data, the role of IT is to provide the right tools in order to have effective ownership and stewardship of data.
- *Metrics and measurement.* Implementing data governance within ones organization is many times difficult to follow-thorough. A problem facing many organization is the fact that it is hard for data governance champions to justify the cost (The information difference, 2010), which evidently, which can be seen from the same study, reveals that companies have a hard time demonstrating the need for DG. However, if approval for embarking on DG has been given the key challenge then becomes to retain the interest for the program and to avoid being cut-back on resources whenever upper-management are looking to make cost-reductions. This then requires the DG program to be able to showcase it success, which Berson and Dubov (2011) argues is achieved by establishing a set of performance metrics that quantifies the organizational success, which in a MDM context is needed in order for the DG to gain executive backing.
- *Effective organization structure.* For companies planning on implementing MDM, this will have significant impact on the enterprise. However, one of the problems with data governance is the lack of follow-through (Loshin, 2009). While one of the critical missions of the MDG function is to form master data related policies, without the proper underlying organizational structure to make them actionable (e.g. through use of data stewards), they provide no added value. Thus, having a robust DG structure in place clearly describing e.g., what the roles are, what decision rights the role holds, and who is accountable for a specific master data, is key in order to meet organizational demands (Radcliffe, 2007).
- *Scalability.* According to (Patton & Wentz, 2010) successful governance initiatives are built on staged implementations, which suggests a governance model that is scalable over time.
- *Data quality.* One of the fundamental reasons for doing MDM is to improve the data quality (Patton & Wentz, 2010), agreed by Berson & Dubov (2011) who argues that one of the objectives of MDG is to ensure that data quality continuously improves. Furthermore, they argue that either new policies be introduced or existing policies augmented to focus on master data quality. Again, this is tightly connected to the *metrics and measurement*, as the success of the data governance function will be evaluated against the metrics that in turn measures the quality of master data.
- *Metadata.* Business terms such as “customer” are used so frequently within different functions in an enterprise which eventually leads to it losing its precise meaning (Loshin, 2009). And after several years give rise to an enterprise lingo confusing for all except for those more senior people who know all of this by heart but with no established framework for extracting this knowledge (Loshin, 2009). The role of the governance organization is then to ensure data is interpretable by developing a formal

approach to how data is documented which enables information tracking – important in many legislative contexts (Khatri & Brown, 2010).

In a white paper presented by Dyché & Nevala (2008) they list ten mistakes not to do when launching a data governance program:

Mistake	Description
Failing to define data governance	Failed data governance initiatives due to the company misinterpreting what it is data governance do and how it will integrate with the organization; many companies do the mistake of using data management and data governance interchangeably which is incorrect (Dyché & Nevala, 2008)
Failing to design data governance	There is no one-fits-all governance model applicable for all organizations. Organizations must tailor their data governance organization and processes to fit into the company context.
Prematurely launching a council	Tied to problem #2, before deciding to launch a council the organization must understand the “what” and the “how” before embarking on the “who” will sponsor and chair the council.
Treating data governance as a project	Data governance is not a one-time-effort but rather a continuously ongoing program. Data change, volumes and increase which requires a structural processes for ensuring data complies with policies.
Ignoring existing steering committees	In organizations with an already established steering committee Dyché & Nevala (2008) consider it foolish not to leverage on existing knowledge. Furthermore, by inviting them to participate in the data governance effort, it will institutionalize data governance as component of corporate policy (Dyché & Nevala, 2008).
Overlooking cultural considerations	Introducing changes to existing processes and behaviors is one of the biggest obstacles with DG. In an example provided by Dyché & Nevala (2008) some corporate cultures stresses consensus and having everyone involved, over accountability. Establishing unambiguous decision rights is critical and how these are assigned according to the authors should not be decided by the cultural norms.
Prematurely pitching data governance	Before soliciting executive sponsorship, pitching in the potential outcomes and assembling working teams, there need to be a framework and vision in place describing how one intended to achieve the vision.

Expecting too much from sponsor	The importance of having executives and management support cannot be emphasized enough. However, it may be the case that sponsors view their role as more of a supporting one, rather than doing the heavy lifting in designing the DG program.
Relying on the big bang	Data governance, especially within the context of MDM, addresses a range of questions such as cross-functional data synchronization, information security, data quality etc. It is common for companies to be tempted to address all of these in a single effort instead of taking an incremental approach to DG i.e. starting small and thinking globally.
Being ill-equipped to execute	Connects back to mistake #4, which is that DG is not a one-time project. As data continuously changes, maintenance and audit are often underestimated. For perceiving DG as valuable, DG must be measured, hence requiring structured management of data.

3.5 Change Management and MDM

MDM has more to do with governance, people and processes rather than technology, managing change is thus key for MDM success (Radcliffe, 2011). Embarking on a MDM program will impose a number of changes within the way the organization which may give rise to resistance due to loss of turf and political interests. Radcliffe (2011) suggests “five elements of change” in the context of MDM which shares many similarities with those found in Kotter’s (2007) widely known eight-step transformation model for successful change.

Imperative. For failed change initiatives there are commonly two patterns that emerge. Firstly, sponsors or advocators for the change assumes the case for conducting the change is clear and hence do not bother to prove it; secondly, not putting enough effort into convincing stakeholders of the change imperative, thus leaving them to believe that it is not real (Radcliffe, 2011).

Leaders. True to all change initiatives is the need to good leaders who understands the imperative of the change and possess needed mandate and personal traits to able to enforce changes over a long period of time, in the face of criticism and resistance (Radcliffe, 2011). This is also agreed by Berson & Dubov (2011) who points to the fact that implementing an MDM is an expensive and lengthy processes where senior commitment in terms of resource, strategy and governance is key. Furthermore, needed executive sponsorship, commitment and ownership is only obtained once sufficient knowledge of the changes in each domain is known; multiple iterations with senior management is therefore to be expected before any projects can be initiated.

Affected agents. Are those individuals or groups affected by the change and who must adapt to the changes imposed, in this case, by the MDM initiative. Radcliffe (2011) argues there to be three main behavior patterns for agents undergoing a change, these are: Change Absorption Thresholds, Change Acceptance Cycle and Change Participation Rates. In essence they address the necessity to assess a company’s current change capacity, to understand whether the organization is equipped to be able to manage the change. Furthermore, as will all changes to a beginning there is a hype around the change which will in time fade away leaving a pessimistic state-of-mind towards the change initiative. Failure to address this, if according to Radcliffe

(2011) where most change initiatives (including MDM) fail, although as this is a repeatable pattern for most change programs it can thus with required capabilities be managed.

Levers and Buoys. Levers consists of those tools used by a change leader in essence levers is the toolkit used by change leaders when employing pressure for obtaining desired results. Examples include, but are not limited to: metrics, positive/negative peer pressure, process alteration, etc. (Radcliffe, 2011). Buoys on the other hand, are employed as stabilizers for affected agents who in time of change may experience uncertainty. Through buoys a positive frame of reference is established assisting individuals or groups through the change; an example could be a rallying point for a widely accepted imperative (Radcliffe, 2011).

These five elements of change, depending on context, must take into account change at three different levels (Radcliffe, 2011) as shown in Figure 6.

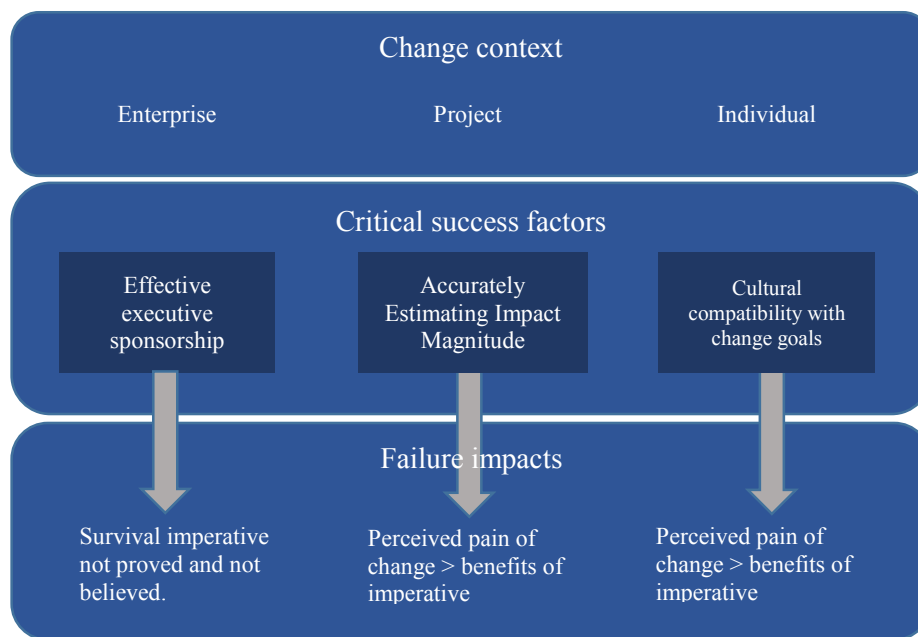


Figure 6 Common Pitfalls of MDM adopted from (Radcliffe, 2011)

At an enterprise level the focus is to ensure that necessary stakeholders at executive level are onboard and understand the imperative of the change. The MDM leadership must thus be able to exhibit a business case which clearly show why this is necessary. This aligns with Kotter's (2007) 8-step model for successfully managing and executing a transformation, for which one of the critical steps – Establishing a sense of urgency" – in an unambiguous manner displaying why the change is critical and what consequences the company may face if not going through with the transformation.

On a project-level, there needs to be accurate assessment of the magnitude of the change and its impact on a personal level. If this is however neglected, there is a risk that the change may be perceived as not worthwhile, leading to passive resistance and half-hearted attempts (Radcliffe, 2011). Then there is change on the individual-level which entails getting rid of obstacles for a change such as, continues use of decommissioned legacy systems, middle managers feeling threatened due to loss of turf, old hierarchies, etc. This is also part of Kotters (2007) eight-step model, where he empathizes the importance of removing obstacles for a change and empowering individuals advocating change by encouraging risk-taking and untraditional thinking.

3.6 Data Quality Management

3.6.1 Data quality management: definition and motivation

The impact of poor quality of data on business is a well-known, as seen from the paper presented by Redman (1998). Poor data quality severely inhibits many strategic business initiatives and makes it difficult if not impossible to generate business value from CRM, BI or any other significant integration effort of data (Friedman, 2007). Data quality is a multi-dimensional concept and comprises subjective perceptions and views of data quality as well as objective measurement based controls (Pipino et al, 2002; Wang & Strong, 1996). Redman (1998) argues that many problems faced by executives have at their root poor data quality; which he further states may lead to consequences such as: organizational mistrust, increased operational costs and ineffective decision making. More often than not IT is blamed for the poor quality, forgetting that data is in fact created and consumed by people outside of IT (Olson, 2002).

In 2002 The Data Warehousing Institute (TDWI) estimated that poor quality on customer data cost U.S businesses 611 billion annually in postage, print and staff overhead¹. Despite this, only a few companies make the connection of bad data quality and increased costs which is instead believed to be due to the lack of monetary measurements of poor data quality (The Information Difference, 2010).

3.6.2 Data Quality, Master Data Governance, Policies and Metrics

Data is the lifeblood upon more and more business decision are based – and consequently with poor data follows poor decision-making making it that the data quality processes should be owned by the business side. Interesting enough, in a study conducted by the Information Difference Company (2010) comprising 257 companies – among the companies planning to implement data governance, they did this with intentions of having a more systematic approach to measuring its data quality.

Many organization may already have data related policies in place, however Berson & Dubov (2011) stresses the importance of either instituting additional or augmenting existing policies to also focus data quality on master data domains. It is important that there is a policy clearly stating whether or not that master data will be managed as a an enterprise resource or not; with support from a higher level principles e.g. a principle stating that all enterprise data should be managed as an corporate asset (Ladley, 2012).

One of the promises of MDM for those companies suffering from their master data being scattered across a number of different systems is the concept of an authoritative master data source. For it to be accepted, quality of data must be must governed cautiously hindering poor quality data to be propagated across the enterprise which may lead to organizational mistrust towards MDM initiative. It is evident that the relation between DG and DQ is symbiotic, as it is the task of DG to ensure data quality is accepted and adapted by the organization (Ladley, 2012). From the Information Difference (2010) study out of 257 companies surveyed 80 had already adopted data governance, while another 39 considered implementing DG prior to embarking on DQ and MDM to be key for a successful MDM initiative. Which again, ties back to one of the ten mistakes – “prematurely pitching data governance” – found in (Dyché & Nevala, 2008) for which authors argue that the DG vision must be clear, while also having a framework established explaining how to systematically approach it.

3.7 Data stewardship

Data stewardship (DS) is crucial for the success of data governance; they enforce policies on data and ensures that changes lies in the best interest of the organization. David Plotkin (2014) describes data stewardship as “the operational aspect to data governance , where each data item should have a designated steward who is responsible for most of the day-to-day managing of the data. Friedman (2007) on the other hand recommends data stewards be appointed to large

¹ <http://download.101com.com/pub/tdwi/Files/DQReport.pdf> (accessed 2014-07-13)

business units, in which they are regarded as subject-matter experts. For the stewardship role to succeed it is important that the business are made aware of the quality deficiencies of its data (in this case master data). To achieve this both Berson & Dubov (2011) and Loshin (2009) argues that the stewards should reside within the business as opposed to IT since they are ultimately the consumers of data and has the best overall knowledge of the impact of poor data. The steward should for the sake of efficiency be positioned close to the point of data creation and consumption (Berson & Dubov, 2011) and to coordinate between these two end-points.

One of the common challenge of data is trying to understand data and its areas of uses, and with no appointed stewards and metadata responsible, individuals or departments (usually IT departments) tend to guess and figure things out by themselves in the light of pressured timeframes (Plotkin, 2014). This is where data stewards come in, and plays a crucial role for MDM implementation. Ted Friedman (2007) describes six responsibilities or characteristics of data stewards as part of a DG program:

- *Subject-matter experts.* Steward(s) responsible for e.g., customer, should be specialist who are in charge of ensuring that customer related data adhered to corporate data quality standards.
- *Guiding as oppose to executing.* Since the steward may have other roles and responsibilities within the organization to attend, hence the steward is responsible for guiding and influence changes rather than taken the role as the “doer”.
- *Focused goals for data quality improvement.* It is important that clearly defined, and measurable, goals exists for the data of interest. Moreover, the data stewards should assist in defining appropriate quality metrics.
- *Accountability.* As the role of being a data steward often is a smaller responsibility of an individuals overall responsibility, consequently, the stewardship role may be deprioritized by the individual or steward group. To prevent this some enterprises may use incentivize stewards through means of additional compensation.
- *Manageable portions of data.* It is important that data stewards are given responsibility for manageable portions of a data entity. The risk otherwise, if provided the sole responsibility for a large data entity is that there may not be enough time to address all deficiencies that may exist.

A common mistake is failing to separate between the ownership and stewardship roles. Where ownership refers to those individuals or groups who have the authority to create and have control over its content (Friedman, 2007; Berson & Dubov, 2011). Whereas, data stewards do not own the data, but instead regarded as data custodians (Friedman, 2007).

It is important that data steward claim accountability for the quality of the data they manage, however this will in turn require sufficient empowerment from the organisation accepting the data stewards as the prime responsible for the quality.

4 Methodology

In this chapter a description for how the thesis was conducted will be given along with justifications for some of the design decisions made.

As presented in the introduction the objective of this thesis is to answer the two posed research questions. With regards to research question 1 (RQ1), in order to understand the shortcomings and what challenges faces NCC in terms of managing their master data in essence it is necessary to understand i) what course of event preceding this thesis had led to the current state of data ii) how is master data currently being manage and lastly iii) how do employees perceive the state of master data. To answer these questions and eventually RQ1 a qualitative approach was taken by interviewing NCC employees from both business and IT. Furthermore, complementing the interview data was informal chats with NCC employees and observing the environment in which many of the employees operates in as it may influence what is said, or not said, during interviews and can help contribute to the authors understanding of the situation.

With the basis from the findings obtained and presented for RQ1, RQ2 is answered by conducting interviews with external consultant to get their point-of-view (considered best-practice) on some of the ways the shortcoming and challenges experienced by NCC can be addressed; as well as comparing this to what is advocated by literature.

4.1 Research Approach

In (Collis & Hussey, 2003) the authors argue that two main research paradigms exists: the *positivist* and *interpretivist (or phenomenological)*, better known as the *quantitative* and *qualitative* research paradigms. The positivistic paradigm is based on the assumption that social science research should be conducted in the same manner as that found in the natural sciences. The positivists therefore argue that social behavior and actions can be explained by logically derived laws which would enable one to anticipate the occurrence of the phenomena (Collis & Hussey, 2003)

There exists three modes of inference *deductive*, *inductive* and *abductive*. The deductive reasoning proceeds from an existing set of theories and laws to generate a hypothesis which can then be tested of its validity through empirical observations (Collis & Hussey, 2003). In contrast to the deductive approach, the inductive reasoning instead assumes there is no preexisting theory; and that theory is empirically formulated by making general inferences from specific instances (Collis & Hussey, 2003). The third inference mode – abduction – combines the strengths from both an inductive and deductive to gain new knowledge (Svennevig, 2001) and continuously reformulating the theory as an increased understanding is obtained.

The research paradigm used in this thesis has been interpretivistic. This approach was considered more suitable compared to the positivistic paradigm as the research seeks to understand the success factors of deploying a master data governance function; which entails interpreting underlying meaning of what is said during the interviews. This is then analyzed together with existing theory in order to identify or verify the existence of any reoccurring problem patterns (or intervening conditions); seeking to understand how or if they are connected to each other, by means of causal relations. Since any interpretations made of the empirical data is evaluated from the interviewer's own reflections and understanding for the situation. The author has been aware of the risk for biased reflections, has thus taken some measures (explained later) to reduce this risk.

The empirical data has been collected and processed inductively with some deductive elements involved thus making the approach an abductive one. This approach was considered favorable as the problem with master data was not known before-hand and instead had to be identified for which unstructured interviews together with observations were employed during the early phases of thesis. The acquired knowledge, from the unstructured interviews where then matched against existing theory to better understand the underlying problems that would later pave the way for a more structured data collection.

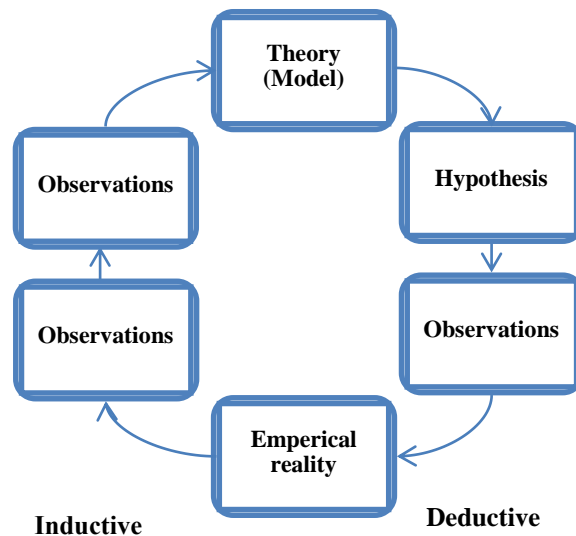


Figure 8. Induction and deduction, adapted from (Wiedersheim-Paul & Eriksson (1991) cited in (Le duc, 2007)

4.2 Research processes

4.2.1 Data collection

When collecting the data two types of data sources have been used: *primary* and *secondary* data. The primary data has been gathered through unstructured and semi-structured interviews. The secondary data has been gathered through scientific journals, white papers, articles, eBooks and regular book – some found online and others from the libraries at the Royal Institute Of Technology (KTH) and Stockholm University and are presented in §3.

Interviews

As previously mentioned both unstructured and semi-structured interviews were used for this project. Distinctive for unstructured interviews format are that questions are not prepared before the interview but are rather formed during the course of the interview. Whereas, semi-structured commonly have questions relating to a pre-defined theme (or themes), although at a level such as enabling the interviewees to speak his/her mind within the frames of the interviews theme. Additionally, since the first part of the thesis (i.e. answering RQ1) is concerned with identifying the shortcomings and challenges in how NCC currently is managing its master data. Therefore constructing the interviews questions in such a manner that negative factors can be extracted from the interview data is a key point.

The purpose of the unstructured interviews was to narrow down the scope of the thesis problem by investigating the problems related to master data. In parallel with the unstructured interviews the literature study phase was started. The combination of unstructured interviews to identify problems and then consulting literature of the matter has proved beneficial to get an firm understanding for the topic of master data management, data governance and also data quality; and how they relate to the situation at NCC.

The output from the pre-study phase (literature study and unstructured interviews) were later used as basis for creating the interview questions used during the semi-structured interviews. A semi-structured interview format was used to understand the problems involved in instituting a MDG function. As one cannot expect to be provided the causes when asked for them directly, since this may not be apparent for the respondent himself. And so, this had to be taken into consideration when designing the interview questions. Furthermore, the possibility to ask follow-up questions for systematically getting to the core of a problem is also why a semi-structured interview format was preferred.

Since a semi-structured interview format was used the questions used in the interviews were centered on the predefined themes deduced from the unstructured interviews. The interviews took place with a number of employees, all of whom in some way, either direct or in-direct, work with master data. To further strengthen the reliability of the results - respondents were selected to represent views and opinions relating to master data from both from a business and IT perspective.

Since an overwhelming majority of literature reports that successful MDM deployments (including the deployment of a master data governance function) resides with organizational challenges (e.g. people, policies, change management etc.) rather than with technology, this justifies the rationale of involving people from both business and IT during the interviews; which were conducted in a face-to-face setting.

Prior to every interview the respondents were, one or three days in advance, sent a short description of the topics of the interview so that they would have the chance to reflect on the topic. Each of the semi-structured interviews were audio recorded and shortly after transcribed.

To ensure that the risk for any biased reflections were held at a minimum a short summary of the transcription, highlighting, in the form of a matrix with problems identified by the author along with its corresponding causal factors and consequences if not addressed, were sent back to the respondent for validation.

Name	Company/Title	Business/IT
Respondent 1	NCC Group	Business
Respondent 2	NCC Housing	IT
Consultant 1	Deloitte Consulting/ Senior manager	N/A
Consultant 2	INNARC/Partner	N/A
Respondent 5	NCC Group	IT
Consultant 3	Accando/ Senior management consultant	N/A
Respondent 7	NCC Group	IT
Respondent 8	NCC Group	IT
Respondent 9	NCC Group	IT
Respondent 10	NCC Construction	Business & IT
Respondent 11	NCC Group	Business & IT
Respondent 12	NCC Group	Business
Respondent 13	NCC Group	IT

Table: List of respondents

4.2.2 Interview question design

4.2.3 Unstructured interviews

As the first set interviews were conducted using an unstructured interview format no detailed questions were prepared for these interviews, instead while reading up on MDM some reoccurring areas within the field was identified and used as a basis for starting of the conversation about master data. The areas included: data quality, data governance (roles, policies, ownership) and processes.

4.2.4 Semi-structured interviews

The purpose of the semi-structured interviews was twofold. Firstly, this would yield an understanding for some of the more non-trivial problems faced by NCC today in establishing a data governance function, along with additional data (e.g. context, actions taken, intervening conditions and outcome). Secondly, the gathered interview data would then form the basis for creating a causal graph in which the identified problems are depicted as well as its casual relations to other identified variables.

As evident, the structure of the interview questions is of paramount importance in order to attain insightful interview data. Therefore the interview questions had to be designed in such a fashion that they would allow question such as *why*, *how*, *what* and *when* to be answered.

The first approach initially considered was that of grounded theory (Collis & Hussey, 2003), which is a robust and powerful method divided up into three steps (open-coding, axial-coding, selective coding) commonly used to form a theory. However, due to its exhaustive nature and the limited time frame full use of these method was considered impractical. Nevertheless, other aspects of the method (open-coding and axial-coding) served as inspiration when designing the method for use in this thesis.

Similar to grounded theory, the method consists of a coding phase, with the difference that the codes (or rather categories) consisted of a pre-defined list of categories frequently mentioned in literature. The list of categories would expand whenever additional categories were encountered while reading through the interview transcripts.

4.2.3 Interview Data analysis

Since the objective is to construct a causal opinion graph from which later conclusions can be drawn in this section the approach, consisting of systematic steps in order to identify the problem categories and potential causal relations among these, will be explained.

This framework draws inspiration from the Grounded Theory (Miles & Huberman, 1994) a widely used qualitative research method employed within the social sciences. It focuses in trying to propose a theory based on data extracted from respondents which are then analyzed in iterations forming the basis for a hypothesis explaining a series of phenomena using both inductive and deductive reasoning.

However due to the time-consuming nature of the grounded theory method, some concepts (namely open-coding and axial coding) of the method are used stand-alone outside the grounded theory method as a basis for creating the causal graphs. Described more in detail later during this section.

There is a difference of opinion regarding the adequacy of using qualitative analysis methods for causal explaining (Maxwell, 2004). With critics arguing that qualitative research methods by themselves cannot establish causal relationships and explanations without verifying through experiments. A view which is not shared by Miles and Huberman (as cited in Maxwell, 2004).

4.2.3.1 Coding

Open coding

Shortly after each interview transcript had been finalized they were individually analyzed with the purpose of reducing the volume of data to that of most relevance to the research questions. As a first reduction open-coding is used where the transcripts are read in a line-by-line manner attaching codes (using annotations) to words and sentences referring to explicitly stated problems in the transcripts, with an undertone of problem and/or concern or mentioned in literature. The codes created were short and concise (preferably consisting of a single word) and entered into the first column of the table depicted in Figure 9.

Axial coding

After the completion of the open-coding the codes are analyzed one by one with objective of identifying the causing factor(s) and consequence(s). By causal factor it is to be interpreted as those factors either through inductive or deductive reasoning that contributes to identify problem identified during the open-coding analysis.

Likewise, the consequences column in this regard is to be interpreted as the outcome of this problem which may or may not be explicitly stated in transcript. For identifying the causal factor(s) while reading the transcript the author continuously had to ask himself the question “What is the preconditions causing this problem?” which proved to be a good way to enable a more creative thinking in cases where the answer was perhaps not readily available in the transcript. If identified in the transcripts annotations were made in the text describing the reasoning which is helpful when retroactively trying to understand the justification for the factor (this was especially useful when analyzing the relations between the variables).

In the same manner as for the analysis for causal factors the question asked for identifying the consequence(s) has been “what is the action of response to this?”. Since understanding why previous initiatives within NCC has been unsuccessful with regards to master data it is

important to understand what actions were taken as result of the causal factors leading to the identified problem (or phenomena).

Numerous iterations analyzing the tables were taken in order to improve the result by example making the inverse statement i.e., if we have statement A -> B then trying B->A, and assess whether this way of looking at it makes more sense. This technique of stating the inverse is argued by (Miles & Huberman, 1994) as a good way of increasing the quality of the conclusions drawn.

Challenge/Shortcomings	Causal factors	Consequences
...
...
...

Figure 9 Structure in which data was entered during coding

4.2.3.2 Consolidation

First stage

As an additional measure of action for increasing the quality of the conclusions is to eliminate any rows present in the table which are not mutually exclusive. This was accomplished by primarily looking at the identified consequences (or any of the other columns) in order to assess whether there are rows that have similar characteristics and thus are candidates for a potential consolidation. In Figure 10 an excerpt exemplifying an instance where a consolidation was made is displayed. Here it could be seen that challenge (1) had similarities with the consequences of challenge (2), furthermore, it could be seen that they also shared the causal factor “IT-driven” also indicating that challenge (1) is in fact given by challenge (2) why these are consolidated (see lower part of Figure 10).

Challenge/Shortcomings	Causal factors	Consequences
Master data initiatives falling short	IT-driven, consensus-driven, unclear vision	Continues use of aging master repository GR
...
...
Failing to reach consensus	Power and turf interests, IT driven	Time consuming decision making, initiative put on hold



Challenge/Shortcomings	Causal factors	Consequences
Successful follow-through	IT-driven, Power and turf interests, unclear vision	Continues use of GR, time consuming decision making, initiative put on hold
...
...

Figure 10 Example of a consolidation during coding

Second stage

After tables had been created and processed during the first stage of consolidation, another stage of consolidation was required, only this time all respondent tables were inserted into one common table for analysis. The purpose of this second stage of consolidation is to ensure that redundant rows are also eliminated on an aggregated level since *stage one* considered only consolidations on a local level (i.e. for each individual table).

Furthermore, since the aggregated level consolidation includes opinions and views stemming from different individuals from different sides in the organization (e.g. IT and business) whenever having consolidating views on these level this factors such as belonging and the context under which these opinions were formed had to be considered.

4.3.2.3 Creating the casual graph

After the coding and consolidation stages had been completed, the next step was to depict the table in a causal network graph. The motivation for constructing network graphs was the increased overview of data as opposed to using tables. By using a network format for displaying the data also enables the author to think both creatively and critically about the results obtained thus far. Network graphs especially proved beneficial to clearly display which wholes in data that there present by which using both in- and deductive reasoning. The use of network graphs are endorsed by (Miles & Huberman, 1994), who argues that they are especially effective for causation analysis.

Following steps were taken when forming the graph:

- 1) The challenges/problems (column 1) identified were translated into oval boxes where
- 2) The casual factors (column 2) identified were also translated into oval boxes from which an arrow is drawn to the overarching problem category this causal factor impacts.
- 3) The consequences (column 3) is treated in similar way as with the

Since the initial graph made consisted of a number a number of variables and connections, countless iterations was taken on the graph, each time with the intention of consolidating logically related nodes into higher-level categories and thereby increasing the readability of the graph.

4.3 Interviews with external expertise

To complement the case study and the literature study, expertise interviews with three external consultants were held. They were selected based on their experience from previous MDM and DG implementations and interviewed to gain a practical point-of-view on how the challenges could be approached i.e. success factors.

The interviews with the consultants did not undergo the same processes as with the semi-structured interviews performed with the NCC employees. As the purpose of the external consultants was to complement the already created causal network graphs with suggestions based in best practice on how some of these challenges could be addressed; as opposed to create an stand-alone graph for respective external consultant interviewed. Therefore, the interview transcripts created for these interviews followed the same structure used when analyzing and creating the table depicted in Figure 10. However, this time the column 1 was populated with the derived categories found in the graph, and with basis from column 1 the transcripts were analyzed.



Deloitte Consulting is a globally renowned consulting firm offering consulting expertise within areas such as tax, strategy & operations and technology. Is currently the world's largest audit and consulting company with more than 200 000 employees globally.



A consulting firm formed by the two former Sandvik enterprise architects Jan Nilsson and Stefan Eriksson. They both have an extensive background within the areas of information management, enterprise architecture and MDM.



A Swedish management consulting company offering services within finance, IT, supply chain management, strategy and operations. Based in five different countries with approximately 1100 employees.

4.4 Method validity and generalization

As defined by (Collis & Hussey, 2003) validity is defined as how well the obtained results explain and answers the studied problem. Since a semi-structured and unstructured interview format is difficult to fully replicate due to parameters contributed by the author, the interview setting and the interviewee. Yet, by using open-ended questions the author is able to obtain a better understanding for any of the underlying problems by choosing to elaborate further on any remarks made by the interviewees. This is evident especially in the case of interviewing the external consultants, as differences in past experiences is likely to be reflected in their answers. Given a longer timeframe for this study, having used one or several case study companies, in addition to NCC, is expected to increase studies generalization.

Additionally, as the author was provided with a seat in an open landscape setting with some of the interviewees seated nearby the author interpretations of the interview material could be biased. Furthermore, other sources of bias are those that would emerge during interviews such situations where author is put in an ethical dilemma (e.g. in the case the interview is in charge of a system or data entity which has been highly criticized by others) or impressions such as change of tone, facial expression and body language, all which may have influenced either the interviewer or interviewee, as interviews are held in a face-to-face seating.

To ensure the correct interpretations of interview data were made, the casual diagram, after it had been compiled by the author, was sent back to the respondent for confirmation.

Lastly, interpretation of others interpretations may reflect the interviews subjective opinions, and thus requires the author to position himself critically to all interpretations made.

5 Empirical findings

During this section the different parts of the empirical study (i) interviews (ii) observations are presented. Using the data found gathered from interviews with employees at NCC and three external consultants from Deloitte, INNARC and Accando. An opinion graph is put forth, highlighting some of the experienced problems (white blobs) with how master data is currently managed at NCC, as well as proposed solutions (orange blobs) to these.

5.1 Interviews

In total 13 interviews were held with 10 of them being with NCC employees and the other 3 with external consultants Consultant 1 (Deloitte Consulting), Consultant 2 (INNARC) and Consultant 3 (Acando). Out of these 10 interviews with NCC employees, the two initial interviews were used to explore some of the problems present at NCC which helped form the semi-structured interview format later used when interviewing the remaining eight NCC employees.

5.1.1 Opinion Graph – Problem areas



Figure 11: Identified problem and challenge areas

During the analysis of the interviews transcript and construction of the tables and causal graphs three major problems areas were derived. In essence the problems experienced could be divided in one of the three problem areas depicted in Figure 11. These are related to each other given the fact that in order to reach a common data definition for each master data NCC must be able to overcome the obstacles that have prevented them to successfully follow-through on a change initiative related to master data. While success in the change initiative is important for reaching a common data definition so is the quality of data which in parts is affected by the plethora of definitions for the master data entities. From interviews it was understood that the absence of a common data definition had made consolidations efforts, with respect to master data, increasingly difficult and troublesome. Furthermore, the amount of reactive work in to assure the correctness and credibility of data means that there is room for improving the efficiency.

5.1.1.1 Common data definition

Since NCC has a long history of having vertically integrated business areas, both domestic and internationally, with area having their own set of IT systems and definitions which in turn given

rise to some of the witnessed informational silos. With little cooperation to reach enterprise-wide definitions on what for example constitutes e.g. a *customer* or *project*; as such, these have been used and defined differently across the enterprise. In terms of master data NCC today have a system - *Generella Registret* (GR) - which is regarded as one of the most critical data hubs in the information architecture. It consists of a collection of databases storing some of the most important master data objects at NCC.

This system was originally not intended to serve as a master system, this has, through the years, brought with it concerns regarding its flexibility and operational efficiency, in the light changing business requirements. And with the new strategic direction of *One NCC*, resulting in one of recent initiatives to consolidate its master data (e.g. customer and financial data) from NCC subsidiaries local master data repositories across the Nordic whom historically have had their own systems and separate set of master data fields. This have however proved to be a cumbersome challenge, as these migrations rarely proceeds smoothly due to the differences in data models, definitions and lacking documentation.

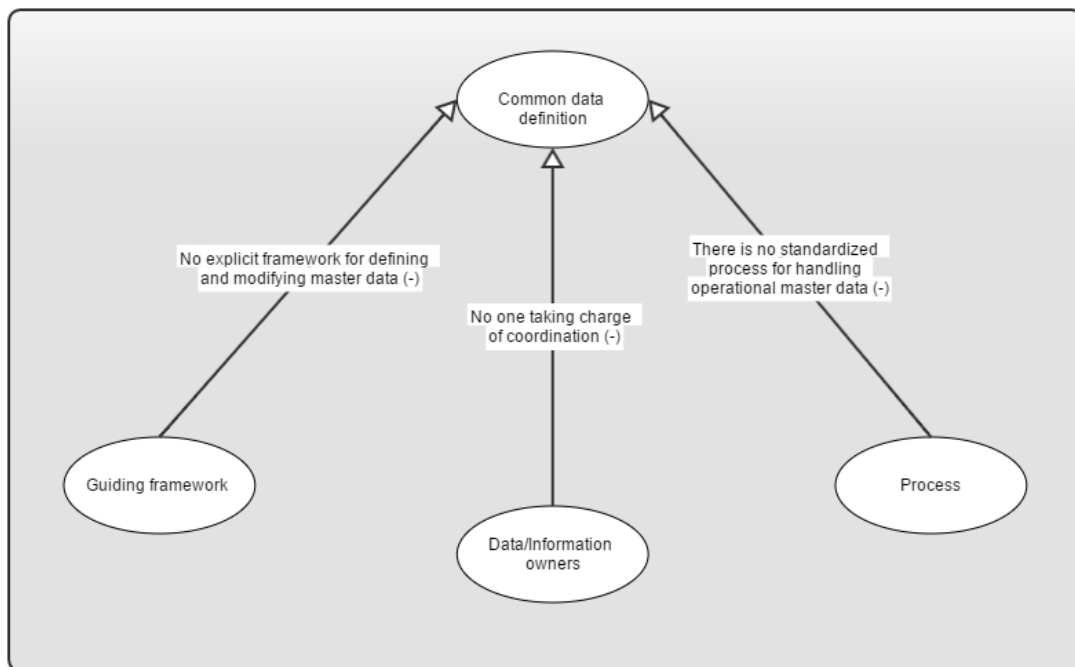


Figure 12: Opinion graph - Common data definition

A majority of the respondents agreed that there were problems with the absence of a common definition for the master data entities. To understand the drivers for this was however less obvious. One of the respondents said:

“There needs to be definition on the concepts one decides upon and routines on how changes are made, so that no one makes this up. There needs to be a decision council.”

This was referring to the *process* of the management of operational data, to which there exists no standardized process today. For financial data there exists processes on how to manage change and keeping track of data that could have direct impact on the financial processes. With regards to the operational data, many of the respondents, and others spoken with informally, did not find this as a surprise as NCC has mainly had a finance focus. This combined with the fact that NCC is a listed company, some of the respondents argue, are why managing financial data tend to be more well established compared to operational data.

The lack of standardized processes for handling master data inquires also leads to confusion with regards to who one should contact when encountering problems or change requests related to master data. Today the business areas either directly contacts the system managers or the

system owner of the systems hosting a particular data entity the data object of interest. This becomes problematic as neither of these two roles have enough insight or understanding of the consequences of accepting changes to master data, which in turn makes it increasingly hard to reject them. In addition, since there is neither documented process nor any documented information/data owners it becomes difficult to direct the inquirer to the right individual or forum. Which in turn gives an aggravated oversight of what fields already exists in a master entity resulting in redundantly stored fields or fields only of interest for one business area.

During interviews respondents were asked, in their own words, to give their definition to master data. Many of the respondents were familiar with the notion of master data, with most of them referring to entities such as customer, project, supplier, etc. Many of the answers however were mostly high-level, and attempts trying to get a better understanding of which criteras they followed to distinguish whether a field was a master or were not successful. Only two respondents admitted that they were missing a framework that would enable them to make the distinction. Which is also confirmed by some remarks made by respondents, some of the master entities, for which they revealed contained attributes only consumed by a specific business area. For this reason, as seen in Figure 12, the node *guiding framework* was identified, indicating that an absence of a guiding framework assisting in deciding whether an attribute is master gives raise to uncertainty and thereby negatively impacts the ability to arrive to *common data definition*.

5.1.1.2 Successful change initiative

The issue of master data (better known in NCC as common concepts) have been addressed in a number of previous initiative. Focus has primarily been more technical, with concerns regarding the aging master hub GR, which is a homegrown application developed during the 90's. Common for these passed initiatives is that they have failed whenever discussions addressing the data mode has emerged. Most recent initiative, part of NCC's finance strategy, initiated roughly 10-12 years, reportedly failed due to difficulties in reaching consensus.

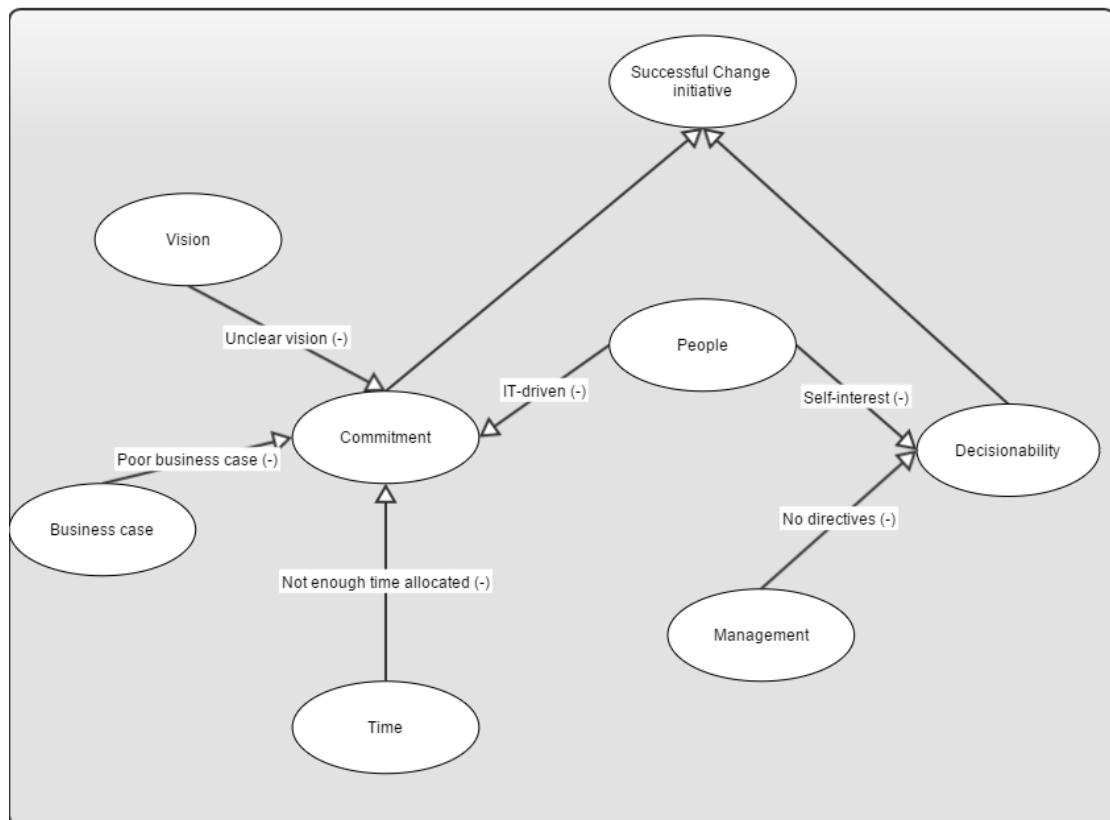


Figure 13: Opinion graph - Successful change initiative

We had an assignment, but it was unclear, not prioritized and there was no “this is the direction” and so it fell through.

Part of the problems, as identified from the interviews, was the lack of follow-through. From the citation it is understood that previous attempts had failed with respect to prioritization and communicating the vision and goal statement. In Figure 13, the arrow drawn from *Vision* to *Commitment* was drawn after it was clear from interviews that if the vision with the initiative is ambiguous this will have a negative impact on the participants' commitment to the cause; which holds true for both operational employees as for upper management.

A number of respondents remarked on the fact that that previous initiatives had been predominantly IT-driven. Which in turn had instilled the perception that designing and deciding on a target common data model lies with IT. In Figure 13 the *People* node is connected to *commitment* where this relation is to display the impact the people has in the commitment of the initiative. In previous initiatives a majority of the participants had come from IT which according to respondents had led to conversations focusing on the technical solution approach as opposed to deciding on what data should be included. Therefore, in Figure 13, the connecting arrow from *people* to *commitment* shows that by involving the wrong type of people (i.e. with regards to background and position) has on the overall commitment to the initiative. This relation was not apparent to the author and required some reasoning taking into account existing nodes and their impacts as well as statements made during interviews.

One of the perhaps most difficult challenges faced by IT is the ability to make good enough business cases as one of the respondents said:

There needs to be a real business case. This does however not imply that there must be any monetary benefits. As long as this is made clear, one must instead try to identify other benefits.

There is strong consensus among the respondents, and with employees spoken to in a more informal setting, that there is a lack of management interest and direction.

Depicted in the figure is also the *time* variable. During interviews three of the respondents told the author that despite being involved in a project still had an obligation to execute ones daily tasks. Adding to which one respondent argued that many times initiatives were put on hold due to poor planning made by the project manager which made it hard for participants to allocate enough time to the project and leaving no margins for drawn out discussions.

..the problem is perhaps that there have not been anyone able to take decisions. There has not been a powerful enough steering and someone that has said “you are responsible for the different concepts we have and the definition of them..”

To a great degree being able to come to an agreement is not a challenge unique for NCC. However, the inclusive culture of NCC, as stated by two respondents has meant that almost “anyone” can have their opinion heard, and since discussions are consensus-driven this presents some challenges reaching any conclusive decisions (as understood from the two excerpts below):

..we got stuck in discussions, we were unable to proceed and make any conclusions saying “this is the way it has to be”. Then there was also a lot of territorial self-interest...

..we in Sweden are very much so, perhaps even more so in NCC, this with consensus and that all should be included and everyone should take part in discussing, which becomes really hard..

Partly based on these above given statements in Figure 13 the *People* node was also considered to impact the node *Decisionability*. There is an agreement among respondents that the inability of reaching consensus in part is entitled to the difference of opinions between business area representatives keeping their own area's interest in mind.

According to respondents the lack of management involvement and providing a sense of direction has had a negative impact on the ability to proceed in projects or to enforce compliance to new processes and systems. Where one example entails a failed attempt of decommissioning an old legacy system, in part due to individuals or groups unwillingness to participate in the change. For which respondents mean that an increased involvement of management in terms of steering the efforts would dissolve many of the obstacles experienced. Therefore, in Figure 13 the node *Management* was drawn out from which a line connecting to *Decisionability* is depicted. Where a lack of involvement (or interest) from management inhibits the decisionability of those driving the efforts since they are not provided with a sense of direction.

5.1.1.3 Data quality

The perceived quality of master data varies somewhat among the employees. When asking a respondent of what they thought characterized good quality of data, following was said:

...hard to say, I have not actually thought about it, but you should only have the information once, right? There should be one truth and not several of the same thing...

The concept of data quality varied somewhat between respondents coming from either IT or the business. Although it was recognized by all respondents that having data quality is essential browsing the intranet for any policies relating to the quality of data were not found. Interviews also showed that the closer one worked with data the more problematic one experienced the data anomalies. After analyzing the interview transcript following shortcomings, along with their belonging data quality dimensions (written in italic), were identified:

- *Incomplete data* – for example fields that supposedly should not be empty can still be left empty. For some data there are field missing that are of great interest.
- *Inconsistency* –for example there was a case where several data entries pointing to a company which according to the data operated in several countries when in fact it only one of those countries existed.
- *Interpretability* - due to the lack of missing documentation of concepts, as well as fields contained within this fields there is an impending probability that, due it being hard to interpret, may result in the fields being used for some purpose it was not originally intended; or that the field is used differently among the business areas. An inherent shortcoming, as a result of the difficulties of interpretation, is that fields within a master data container (e.g. project or customer) although their name differs are used for the same purposes, consequently meaning that there are redundant fields in the objects.
- *Reliability* – the trustworthiness of master data varies somewhat, in addition to the deficiencies in data quality, the consensus is that data warehouses and BI tools may in fact consume incorrect or dirty data.

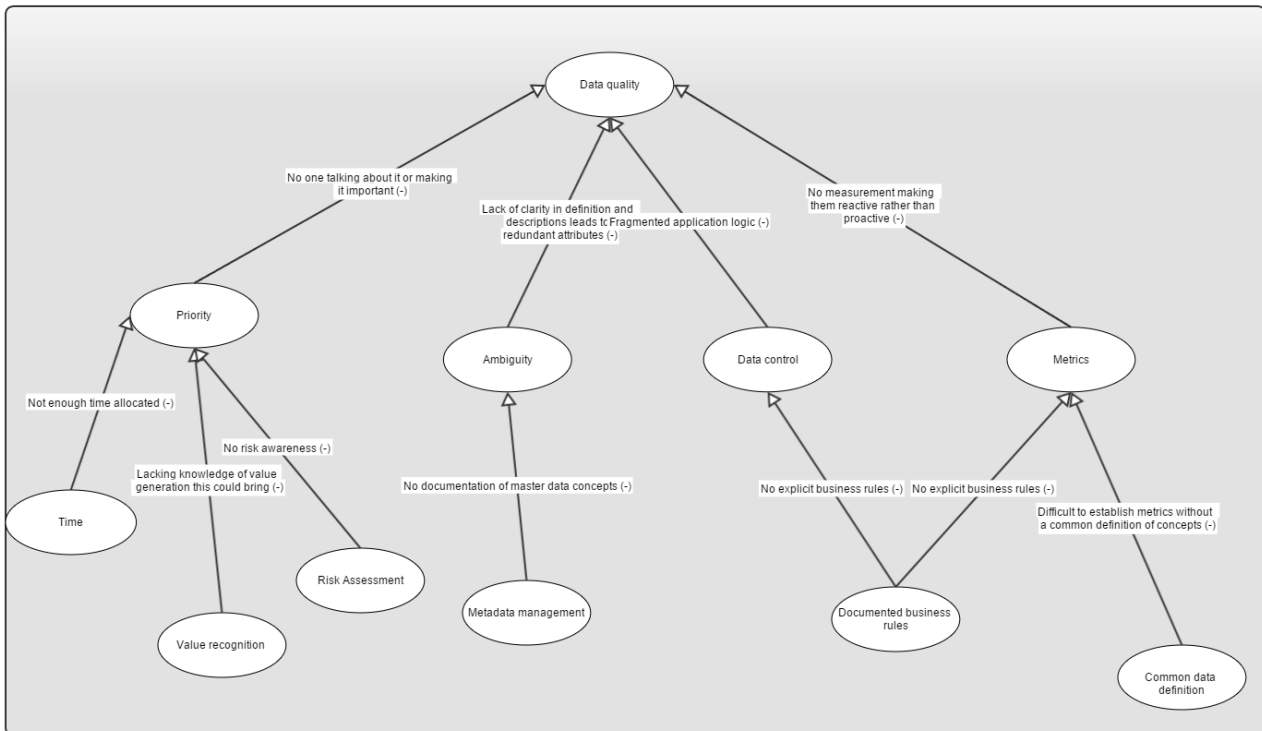


Figure 14: Opinion graph - Data quality

The problem of data quality was quickly realized for the author as it became apparent that there existed no explicit documentation around processes for ensuring its quality of data (especially master data). Today most of the quality work consists of data cleansing i.e., dated fields are identified and removed. When asking respondents of the frequency of these cleansing activities in GR, some differences in answers were obtained where one estimated this to roughly once a year while another gave the indication that this was made on a case by case basis. Where in most cases as a response to a problem raised by those working against the data (e.g. developers or other consumers of data). However, due to the complexity in system integrations and undocumented system dependencies, cleansing efforts, at times, were difficult to complete as whole. Which in turn meant that some problems in data could possibly be left unnoticed or unsolved.

As seen in Figure 14, *Priority* is seen to affect *data quality*, this relation was drawn after taking into considerations that a majority of the respondents admitted that they were aware of some of the data quality shortcomings. Yet, this problems had received little or no action after which the author concluded that this was related to prioritization, also taking the support from following remark made by one of the respondents:

..I think it's a really important question, but then you must make it important, if no one talks about it, if no one explains why this is a problem, why should anyone care to take care of it?..

This was also supported by another respondent who argued that as long as current management of master data (including GR) do not cause interruption to business, or, is not brought up among business leaders as an important aspects to address, current ways of working is likely to remain.

An overwhelming majority of respondents agreed that the importance of improving the quality is not recognized by upper management. Where some respondents believe this is due to a lack of understanding for some of the challenges and *risks* this give rise to in the day-to-day operations; thereby fails to see the benefits (*value recognition*) in investing money to address this. Hence, as can be depicted in Figure 14 is direct relation from the node *value recognition* to *priority* where a low value recognition impacts the prioritization negatively as managers or other stakeholders fail to see the value and the necessity to invest in quality enhancement processes

The shortcomings related to the insufficient descriptions of master concepts and its fields have, in connection with the One NCC direction, become evident. Where one respondent contemplated the fact that there are no documented routines regarding how metadata should be managed, which in turn makes data increasingly hard to interpret and ultimately consolidate. Where another respondent highlighted the potential danger in this, which could manifest itself through unintended use of data.

..we have put in a lot of hours to manage the data anomalies; there can be a value that must not be null, but apparently there occurrences where the value is null. We must set up controls at our end to compensate for the shortcomings in GR..

There appears to be no robust *controls* prohibiting non-valid data to be entered into the master data systems, there are, as exemplified earlier a multitude of shortages with regards to the quality. At the same time as there are no documented business rules or policies to help prevent bad data to be inserted into the fields. Further complicating the establishment of data controls is the fact that the application logic for creating data instances may be done from multiple places each with varying level of control. As an example, due to the absence of comprehensive data controls, there are a number of faulty data entries which not only give rise to confusion but also takes up considerable time having to ensure data is correct. One respondent told the author that they needed to put up controls on their end to verify that the consumed master data was indeed correct, something that should be made at the time of creation.

When asking respondents on how they followed up on the quality of data or had any way of quantifying the impact the current state of data creates. All respondents told the author that they had no way of knowing this expect during the data cleansing where one would get an indication of how much data had been corrected or removed as part of the annual cleansing efforts. Noteworthy was remarks made by respondents saying that it is not very uncommon that invoices get sent out to the wrong addresses or with the wrong information which is as well not measures for following up. Again, respondents agreed that there is still some maturing to be made. Some respondents had observed the fact that there seemed to be no sense of understanding for the scale of the quality problems or what consequences these have had in other parts of the organization. Where one respondent stated that the consequence is that a lot of time is being spent trying to understand and solve data related problems which ideally would be spent on other more urgent areas.

In Figure 14 the relation between the variables *Documented business rules* and *Metrics* were not explicitly stated by respondents, but instead inferred through the reasoning that in order to establish metrics e.g., measuring the number of none-compliant data entries, a precondition is to have business rules defined and explicitly documented; out from which metrics are then derived since the expected outcome is now given.

5.1.2 Opinion - Suggested Solutions

During this section the analysis of the data gathered from interviews with the external consultants is presented. To be able to distinguish between the nodes derived from interviews with NCC employees and those derived from interviews with the consultants, the latter ones are filled with an orange colour.

In the subsequence graphs the consultant nodes are interchangeably either directly connected to a challenge/shortcoming or connected to a causal node. In former case, this is to highlight that a particular suggestion is not targeted towards a specific causal node but as an additional consideration that could help improve a shortcoming. Of course, it may still be the case that it has a positive influence on a causal factor which in then described in the graph-explaining text. In the latter case the suggestion is targeted to a specific causal node otherwise similar as with the former case.

5.1.2.1 Common data definition

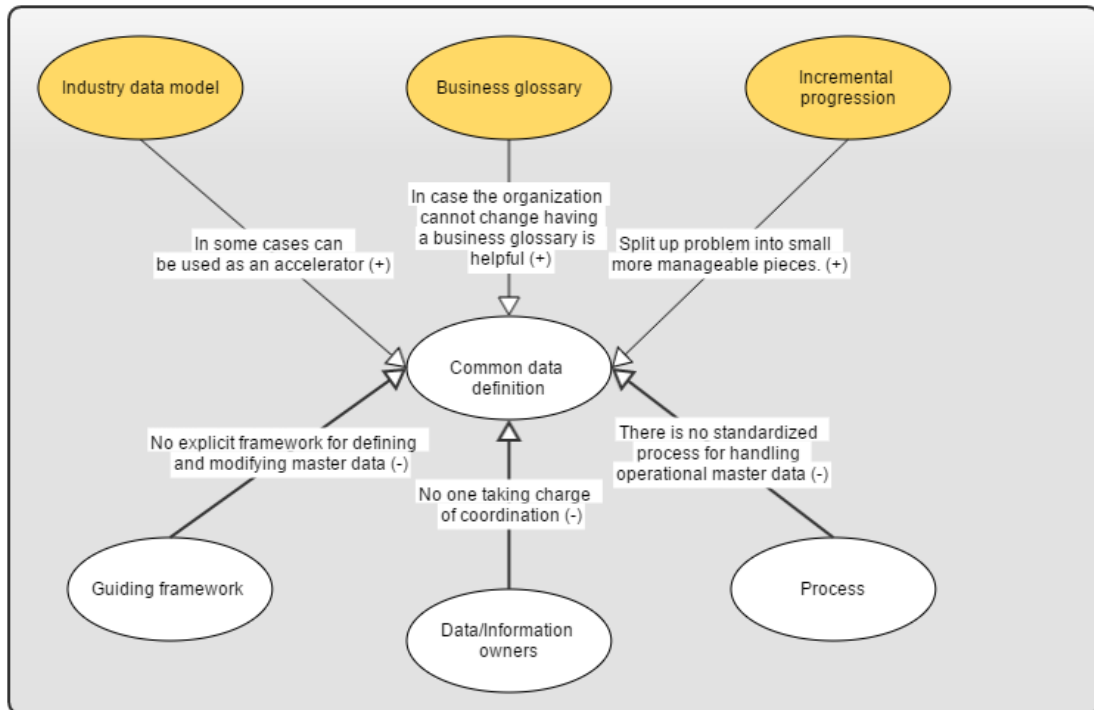


Figure 15. Common data definition - proposed solutions

In Figure 15 the orange colored ovals denotes the views extracted from interviews with external consultants considered best practice on how some of the nodes negatively impacting nodes, identified in previous sections can be alleviated. In the text following this a description for how these variables can help alleviate these problems will be explained.

A remark made by one of the respondents while discussing approaches to reach a common master data definition, was to start from scratch using a blank paper and define the objects one an attribute at a time. However, as this is a large non-trivial endeavor and one that leaps the risk of becoming an overly time-consuming activity and thus not feasible. Instead of starting off with a blank slate another recommended approach is to take an industry data model which typically covers most of the necessary attributes (Consultant 1). By using a predefined industry model as a baseline, discussion are accelerated, as business unit representatives are able to recognize that they share a lot more concepts amongst each other than they realized. Which have previously been impossible due to years of business area specific lingo (Consultant 1).

In the context of common data concepts, and especially when there is several countries involved having gatekeepers who are responsible for data and who decides about changes is key factor to be able to prevent redundant attributes to be augment the data models (Consultant 3). See Figure 16 illustrating the processes as derived from the interview. Since changes on master data entities in an organization such as NCC surely will affect multiple systems, having a council with representatives understanding the impact of the change from different perspectives is important (Consultant 3).

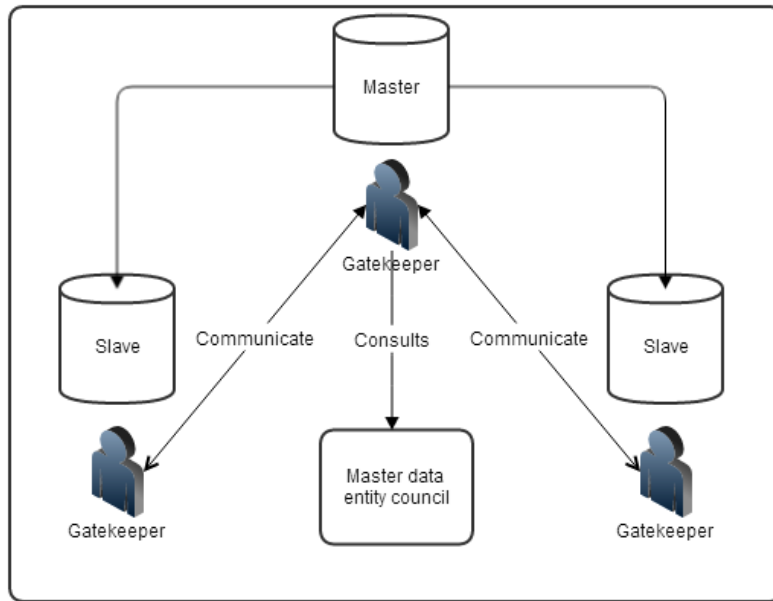


Figure 16: Governing Structure as derived from interview with Consultant 3

A common pitfall experienced by organizations is that they try doing everything at once; a typical example could be that an organization decides it want to define the information object customer. The problem with this according to Consultant 1 is that the object, most of the times, is too complex and complicated for the organization to be able to reach consensus. A success factor in this regards is to get a seemingly big information object, such as customer, and into something that is smaller and more manageable hence pursue more incremental progression.

5.1.2.2 Successful Change Initiative

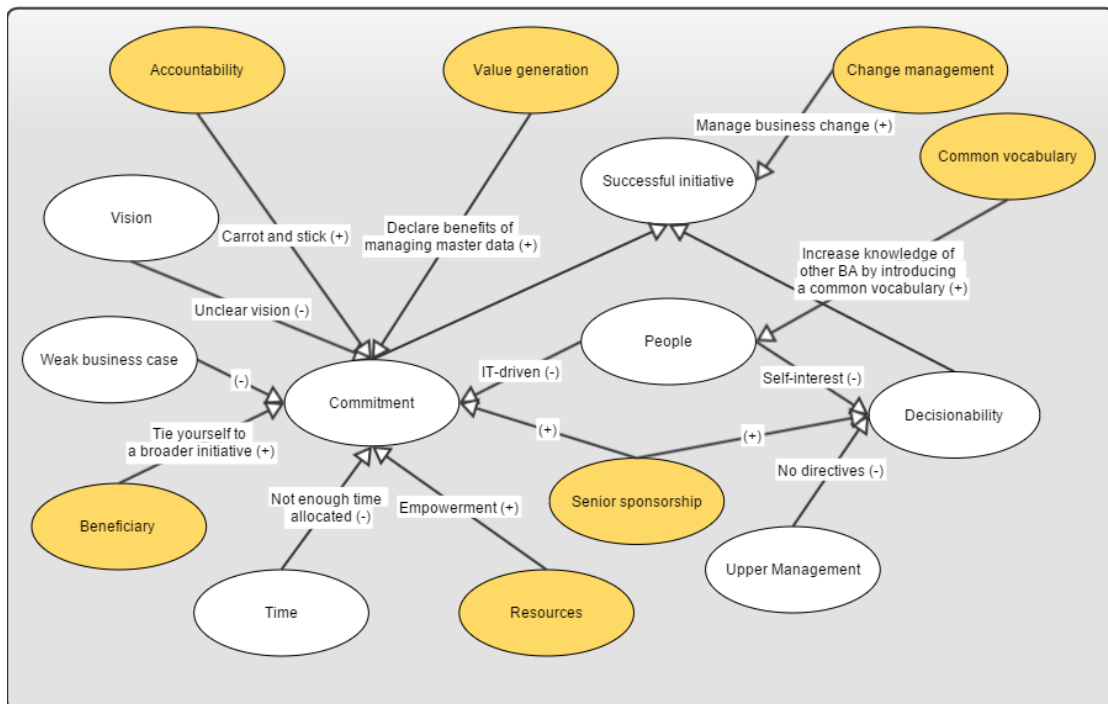


Figure 17: Opinion graph (solutions) - Successful initiative

From interviews it was understood that previous initiatives to address the master data system had failed due to problems of arriving at a common view of what defines e.g. customer. Consultant 1 stresses the importance of tying the initiative project to a broader initiative for where having good data is critical e.g. for ERP, DWH or BI initiatives. In essence identifying your data *beneficiaries* and using them to introduce data governance rather than embarking on a data governance program just for the sake of doing data governance, which according to Consultant 1 is deemed to fail.

The problem of self-interest among participants during change initiatives have had negative effects on the projects in the form of drawn-out discussions concerning details of each representatives own business area. For which the consultants agree that there are no simple answer to overcome this, however one respondent suggests that one should try to introduce a *Common vocabulary*. According to the consultant this requires understand the requirements of each business area, with emphasis on talking with them in their own language. Then this is finally condensed into a set of terminology which all parties can then understand. The other consultants agrees with this, arguing that many times this self-interest emerges due to employees' unwillingness to adapt to new routines which are often met with the phrase "this is how it has always been done".

According to Consultant 2 there is likely an owner for the target state, with whom engaging in discussion with this owner with basis from their perspective is important. Having a gross list of how things look today and in consultation with the owner deciding which of the irregularities on local level related to the information that must be bargained away. In those projects where this approach were not successful, was that in order to have the business areas buy-in to the transformation they allowed local variants to take shape, which in the project one of the consultants were involved with turned out to be a bad decision as it gave rise to even greater problem further down the road.

Consultants agree that having the necessary *resources* in terms of budget, decision mandate is essential in giving data governance teeth so that not only stakeholders stick through the transformation but also project members and other driving the project, in order to empower them to address the problems.

Having senior sponsorship is key whenever embarking on an enterprise initiative such as MDM. Further Consultant 2 explains that it is equally important, and especially so with regards to data governance, that the project steering group be consisting of employees from the business; as they naturally possess a better understanding for how a change will impact the various integration areas in the organization. Likely there is going to be contracts tied to master data, and people wanting to change and complement the information in the master data, which in any case must be escalated up to a steering group which takes into account all the stakeholders' ability to consume and produce this data.

Depicted in Figure 17 is the node *accountability*, as suggested by the author, to strengthen the level of commitment from stakeholders. Although this relation was not explicitly stated during interviews, remarks made by both NCC respondents and external consultants suggested that individuals may feel more encouraged to endure (commit) to a change if they feel incentivized. One consultant brings up the concept of "carrot and stick" by which commitment or non-commitment to a change is either rewarded (carrot) or penalized (stick).

When asked of what it would require from NCCs part to institute accountability for some of the master data entities, following was said:

"..I think it is a journey before arriving at it. It is not easily done and will require some effort. Then again, I do not know, we have a lot dimensions that first have to be taken into consideration.."

From the above given excerpt, the respondent refers to the dimensions as a part of the challenge for introducing accountability. With focus primarily on the two dimensions:

- Geographical – How will accountability be set up taken into account that NCC is present in a number of countries across the Nordic region?

-
- Business area – How will the accountability for master data entities be divided, taking into account that master data as a shared asset is consumed and shared by multiple business area?

According to one of the consultants, this is a common challenge for enterprises embarking on MDM, which further emphasizes the importance of DG, as one would typically expect this to be addressed by a DG function (Consultant 1). For a DG function overseeing common data definitions (as is the case with master data) people from different parts of the organization are typically involved, for which the consultant argued is why you need to manage the organizational roles and responsibilities in a matrix format. Organizations finding this the most challenging are those who do not fully understand the concept of matrix management, or try to be more process-oriented where adding DG is an incremental process.

5.1.2.3 Data quality

The issue of prioritizing data quality was seen to be affected by the two nodes *capacity* and *value recognition*. All three consultants emphasized on, with regards to data quality, the importance of having buy-in from upper management and downwards thereby making them understand the benefits of managing the quality. Consultant 2 admits that quantifying the monetary benefits of improved DQ is difficult, although looking at from another perspective one might ask what the current and potential consequences in the future are if not addressed. In short, being able to demonstrate how poor data quality prevents or delays the enterprise from reaching its business objectives (Consultant 2) which in turn requires that both management and operative staff receive the appropriate *training*.

Establishing accountability for the critical data is one way to ensure that it will be managed properly; at one of the clients Consultant 1 is currently working for a client with data quality controls, and in which case a flag is raised if a data anomaly is encountered. The responsible data owner then has 30 days to address the problem otherwise a charge is taken from the cost center belonging to the data owner.

Since there are no clear guidelines at NCC regarding how changes to the data model should be processed, as stated earlier in this section, this has hampered the organization's ability to get an overview of what attributes currently reside in the data model and more importantly what they mean and how they are used. One of the reasons organizations are adding redundant attributes is due to them failing to discover that a particular attribute already exists because of country-specific denotations. One approach for improving the ability to reuse and increasing the understanding for the data is to introduce a synonym list for each master data object where the respective attribute is listed and along with a list of synonyms used across the organization (Consultant 1).

Gathering and understanding the requirements from different stakeholders for their respective part in the information as early as possible is important in order to address them at point of creation (Consultant 2). This view is partly shared by Consultant 3 who also fills in that it is important to understand who the critical transactions with master data are so that appropriate focus is put on the right objects and attributes.

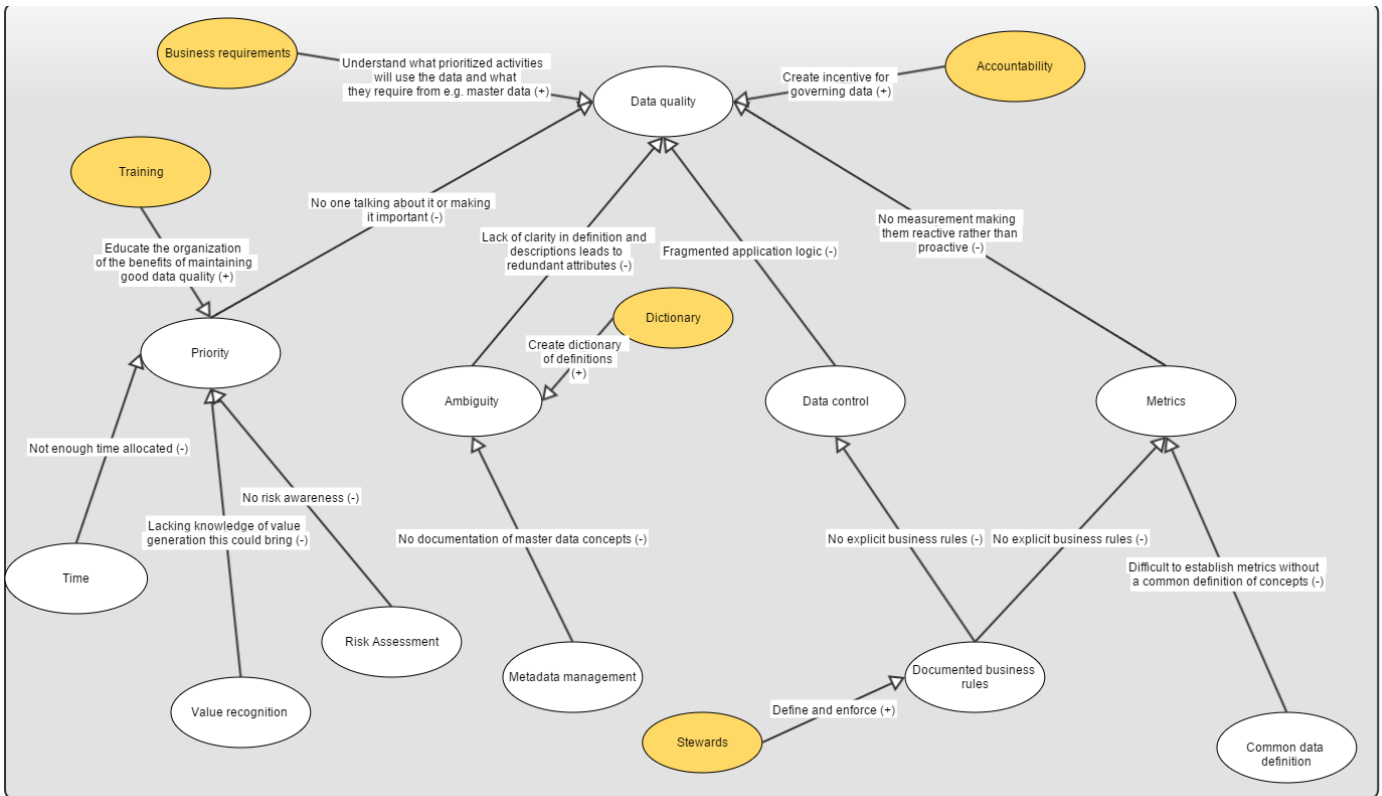


Figure 18: Opinion graph (solution) - Data quality

There is currently no established data quality metrics, partly since it has been unclear what to measure and why. Deriving appropriate metrics based on the strategic objectives set out by the organization and successively breaking them down into measurable metrics is one such approach.

The causal relation between the two nodes *steward* and *business rules* were identified during the analysis taking into account, among others, following excerpts were made by consultants:

- 1) *One thing a lot of companies set up is data governance boards, so data governance boards are there to make decisions on data*
- 2) *..that is what master data management and data governance is all about. It is defining what the rules are and then putting in place controls to make sure that the road rules are followed*
- 3) *..policies and guidelines are for ensuring what is said in the strategy or vision, which you would expect business backing on...when then arriving down to IT and technology additional policies and guidelines can be formulated however pointing to a number of superordinate principles and guideline so that you get an architecture for your principles, policies and guidelines..*

In short, having a master data governance council defining the high level principles with anchoring in the business strategy. These high level principles are then further concretized in terms of policies and business rules. On the back of this, appropriate metrics can be formulated and thus enabling follow-up on how well the enterprise is doing.

5.2 Observation

One of the first things that struck the author during the thesis period at NCC was the hardness of understanding what people managed the systems, what processes existed as well as who owned them (focusing primarily on the group-wide systems and master data). Speaking with system managers it was clear that this was one aspect that took considerable time to understand, and at best system manager keep a contact list that they have made themselves to know whom to contact in case of questions or problem reporting. One respondent were asked if they knew who

they would contact in case of a problem; whereupon this person said following: “yes now with experience, however it took me a couple of years before I had a good enough overview of the people to turn to”.

The poor documentation was also followed by the observation by the author that the number of experts surrounding the master data system GR was sparse. There are currently two persons that have a in depth understanding for the system (with one being a consultant) which presents NCC which, as some argued, presents NCC with a huge risk if competence is not documented or shared in a proper manner.

There seems to be awareness to many of the shortcomings related to data, especially master data. With most arguing that there is an absence of procedures and governance that makes this hard. Furthermore, there seems to be a differing perception to the problems that current management of data can cause. Those employees that tend to understand the connection between complex point to point integration and poor data quality are also the same persons that during interviews and discussions sensed an increasing urgency to address these matters. Although there is passivity, many of the respondents have known of these issues for years but have not acted on them.

At times material located at the intranet e.g. the change management process presented provided by PMO was unheard of which led one to believe that there limited knowledge of what was being posted there.

The official group language is English; however problems arise when discussing master data fields as there is no agreed on guideline whether these are referred to using the English nomenclature. Some mention that this may lead to miscommunication which may affect quality of data. As an example, a concept widely used at NCC is that of *accounting unit*, this concept is however not found in the dictionary over common concept. Instead the Swedish counterpart *kostnadsställe* is found in the dictionary along with its English translation being *Cost Centre* instead of the expected *Accounting Unit*.

There is an extensive use of consultants at NCC today, as described by an employee: whenever there is problem that must be solved typically a large number of consultants are brought in to “put out the fire”. This is however done with no proper steering of the consultants who with their carte blanche, tend to be contributing factor for the failure of IT change initiatives. One respondent, during one of the informal chats told the author that it was important that frameworks be well anchored into the organization with respect to culture and structure.

Another observation made by the author was that some consultants had been with the organization for several years and had in depth knowledge of the IT landscape and its much complex integration. Much admitted that documentation was scarce as there are to established and agreed on routines for how this is should be done.

IT at NCC, and probably for the majority of the construction companies, is still at its maturing stages and the roles such as IT architects and IT strategist although appointed have not worked with tasks in line with their roles. As NCC is one of the leading construction companies in the Nordic region, the concept of architecture is well established in the context of construction but less when it comes to IT.

Furthermore, during the authors time at NCC it was observed that many of the individuals spoken with in both formal (interview) and informal settings felt unease to speak of some of the problems or to target any direct criticism towards, in this case, the business and upper management. Which suggested that the role of IT, as understood from those observed and spoken with, were still considered not as important and with pressing demands for cost reductions employees may feel discouraged to take action or highlight the shortcomings within master data.

6 Discussion

6.1 “What shortcomings and challenges are there in the way master data is managed at NCC today and what causes them?”

6.1.1 Common data definition

In (White, 2010) the author stresses the imperative of having a common master data definition in order to have a successful MDM program. Judging by the past initiatives, both successful and not, that have been launched e.g., the purchase portal (swedish: inköpsportalen) and the one finance project, there is growing understanding for the importance of having structured data with processes for ensuring its quality. The node *Guiding framework* in Figure 12 was considered to have a high impact on the ability to reach a common data definition. Many of the respondents did not explicitly refer to the use of a guiding framework as a cause for not succeeding reaching a common data definition, although a majority of NCC respondents from both business and IT confirmed that there was no stringent business rule that could be references in order to distinguish which master data attributes that would be “mastered” or not. Literature mentioning the use of a guiding framework is scarce, however in (Wolter & Haselden, 2006) this is highlighted where they suggest a set of dimensions against which master data can be discovered. Remarks made by a few respondents referring to the use of business specific attributes contained in some of the master data entities is however supported by Wolter & Haselden (2006) since deciding on what should be mastered can not be decided on the back of a set of criteria, instead they argue that when deciding on what should be mastered this decision should be made against the context of the business needs. Which in part supports the current processes within NCC where this decision mandate is delegated to respective business area. As an example, NCC Construction Sweden which stands for the largest share of the groups revenues² suggests that it is likely that certain attributes that may not be consumed by other areas could still become “mastered”.

Among the identified causal nodes impacting the ability of reaching a common data definition, the node *Data/Information owners* stood out. Nearly all respondents agreed that many of the experienced problems with data and the prior success with initiatives targeting the master data is related to an absence of a data owner that have mandat to act on the matters related to a specific master data entity. Instead of having the system owner making the decision to whether accept or reject an extension of the data model this is a decision that must be taken jointly with affected stakeholders. A good example is the “customer council” consisting of representative controllers from each business areas which collectively assess the feasibility of a change to the charts of accounts. This coalition resemble what is known in literature as a “Center Of Excellence” (CoE, for short) where individuals with a firm understanding for the implications a change are consulted.

For fundamental master entity types such as project and supplier there are no equivalent forum which is noteworthy and from which I conclude that these are not perceived as critical as the customer entity. In essence, by assigning ownership to the master entities we avoid having business areas or individuals ignoring perceiving the data anomalies or change requests related to a master entity as not their responsibility and thereby instilling a sense of accountability.

² <http://www.ncc.se/press/nyckeltal-och-snabbfakta/>

6.1.2 Successful Change Initiative

“Most MDM-related initiatives that do run into problems and stall are due to governance and organizational, not technology, issues” (Radcliffe, 2007).

The above given citation in a brief way captures the current problem residing in NCC. Attempts trying to replace GR has failed partly I believe due to the focus on replacing the technology and thereby underestimating the efforts involved trying also to redesign the information and data model. Where respondents that had been involved in previous initiatives addressing the master data reported that they had been unable to proceed once they had started to discuss the information/data model. In essence the two identified nodes playing a paramount role in the chances of succeeding with MDM is *commitment* and *decisionability*. With regards to commitment the causal node *business case* was by respondents frequently referred to as a “show stopper” meaning that they thought it was difficult to create a business case for embarking into MDM. This is likely connected to the other node *vision* where having a clear vision and objective of what the initiative will help to solve. Understanding what the business problem is that the initiative is helping to solve is important to have this well understood which is a prerequisite in order to have a solid business case. Amongst the majority of the NCC employees interviewed, there seems to be an understanding for some of the problem areas with the common denominator being the need for improved master data. There seems however to be some disconnection for how employees think it should be e.g., do management want an increased transparency in data for use in various analytical tools, or is there in fact a good reason for not having it. This is why communication is an important aspect that did were not highlighted during interviews. While some respondents complained that they are not good enough when communicating, this was not agreed by others who did not see the communication across business areas as necessary since this should be coordinated in group level.

MDM is not all about technology, although great challenges may lie with implementation of it, the perception is that these matters are often time easier solved than those of organizational and governance related. At NCC previous initiatives requiring that master data be addressed, had then consisted of mostly IT, with the consequences that discussions rather than focusing on the information more often than not started focusing on the technical solution as opposed to understanding what information should be included in the data model. Since the business are the ultimate consumer of data, these discussions should be proceeded with basis from the business problem where IT are consulted to assess the feasibility of the solution. Although IT are the ones that are on the frontline when coming to realizing the opportunities and experiencing the implications poor data has on a day-to-day basis, it is important that they include the business and make them take ownership and drive the effort. Otherwise this runs the risk of being perceived as yet another IT project which, to some degree, have been the case for previous attempts at NCC with respect to the remark made by a NCC employee, saying that previous initiatives had mostly consisted of people from IT; thus leading discussions to be side-tracked

6.1.3 Data quality

There are no established or agreed upon data quality dimensions in NCC which is believed to follow from the lack of maturity of data management in NCC. During interviews it was revealed that many of the deficiencies in NCCs master data had existed for some time without having received the sufficient attention to eliminate them. From interviews it was apparent that the matter of data quality where not a fully understood and prioritized area, where one respondent made stated that this was not a prioritized area simply because no one speaks about, and therefore is not viewed as a problem. Redman (1998) explains that poor data quality can have a negative impact on the organization in the form of increased organizational mistrust and poorer basis for decision making as data driven decision making is only as good as the data they are based on.

In fact one respondent claimed that due to previous encounters with poor quality data (including the data stored in GR) one always makes the assumption that data is unreliable which results in extra efforts having to verify the quality of data and setting up controls on the data consuming end to disregard faulty data.

Surprisingly enough however not everyone shared this perception of the data quality, instead there were some who argued that data have to fill its purpose if there are no complains about it. While it is true that data quality is measured against its “fitness for use”, however if they want to have, for example, improved decision making the data would perhaps not suffice. Again, this comes back to the importance of having a clear vision and goal statement of the initiative which must be clearly communicated out to the rest of the organization.

Another causal node impacting the data quality is *ambiguity*. In the context of MDM this problem emerges as a result of lacking metadata management (Loshin, 2011; White, 2010). Loshin (2011) argues that the uses of business terms and words varies between organizational veterans (i.e. those that have been with the organization for sometime) and everyone else. Previous attempts of having a centralized dictionary of definitions had been set up. However, some respondents argued that this was too “exhaustive” and did not serve its purpose and is no longer maintained.

In connection to the consolidation initiatives that have been initiated this lack of metadata has become apparent where one respondent contemplated that a lot of critical information with regards to the data models and data and system architecture that lies within a few individuals which could prove a real problem if any of them were to leave.

Data control was identified as another causal factor impacting the level of data quality.

Although some respondents were indifferent to whether this did in fact have a negative impact on the quality of data, others thought this as evident given some of the incomplete and incorrect data records encountered in GR. The reason why there seems to be lack of data controls that prohibits incomplete data records to be created or preventing them from entering faulty data is believed to be attributed to the lack of *documented business rules*. Where one respondent had encountered records with values that were supposedly not allowed to be empty, that were created with these values set to nothing (or Null).

The causal node *metric* is also seen to be affected by the lack of documented business rules. The use of metrics is a useful way of getting an objective view of the current state of data quality (Radcliffe, 2007). None of the respondents knew of any metrics used to monitor the data, which can also explain why there are faulty data records that have persisted despite the annual cleansing efforts focusing only on the customer entity. Without metrics it will be hard to prove the benefits of having MDM, which in an IT maturing organization such as NCC will play a paramount role.

All respondents agreed that the lack of data quality processes is the first step to have a more quality aware data management practice. However, as one respondent highlighted perhaps there are data entities whose quality is more important to monitor than others. This is not uncommon and in NCC some entities that have stood out during the thesis have been the customer, chart of accounts, supplier and project.

6.2 “With regards to the identified shortcomings and challenges which are the success factors enabling NCC to improve their management of master data?”

From observations and interviews, it is clear that many of the challenges currently facing NCC with regards to master data, at its core, has to do with the lack of proper roles, guiding policies and mechanisms one would normally expect to be institutionalized by a governance organization. Where one respondent argued that previous initiatives had struggled to come to any conclusive decision due to the lack of involvement from both business leaders and top management to assist in the effort by providing direction and/or articulate requirements set on data e.g. for reporting.

As master data is a shared enterprise asset, involving multiple business areas across the enterprise, and hence an array of different stakeholder views and interests – without proper governance it is difficult, if not impossible, to reach consensus and in turn have a gradual increase of the MDM scope in the long-term. With regards to the management of master data within NCC, there exists no chart of roles and their areas of responsibility. Which according to respondents had caused many obstacles and prevented them from going forward to from a common data model, as well as, escalating some of the encountered data issues to the responsible individual. This in contrast to the technical aspect of MDM, where prior to the initiation of this thesis the increasing complexity in NCC's IT landscape had resulted in the establishment of an EA practice. From which the author concluded that there tend to be a better understanding for many of the technical challenges in NCC as opposed to governance and organizational ones with regards to master data.

There is currently no appointed governance organization set in place to oversee the management of data, which in turn have meant that many of the issues related to, in this case, master data are acted upon on a case by case basis. This in turn promotes a more reactive rather than proactive work to prevent the chances of poor data entering the repositories. To this end, revising and reinforcing controls placed at point of entry will be an important task ahead, and one that according to literature as “best-practice” need to be addressed by a potential governance function at NCC.

6.2.1 Data quality

The imperative of maintaining high data quality is evident, since it is the lifeblood of an enterprise. However both the literature and empirical study demonstrating otherwise, it stands clear that the experienced quality of data may vary depending on who one asks, and also taken into account that data quality in itself is a multi-faceted concept where respondents answers showed that they measured data quality against different dimensions where the most prominent ones being correctness, consistency, redundancy and timeliness. These dimensions in turn reflects the needs and experiences of that stakeholder (Pipino, et al., 2002). The impact of poor quality is highlighted in literature and where Marsh (2004) mentions delay in system developments to be one notable pitfall.

From Figure 14 one of the reasons for poor data was the lack of priority. Although there are some cleaning activities to master data such as customers, on an annual basis, these efforts sometimes fall short due to inadequate tools and the vast number of dependencies amongst systems making it hard to predict the consequences of removing some of the data. With respect to priority Redman (1998) suggests that one of the first steps that must be taken is to raise an awareness for the problem and its impact. To this end, educating employees, on all levels, dependent on good quality data is essential. By demonstrating how improved data quality enables them to be more efficient (Consultant 1). By instituting data quality policies and metrics which are then followed up by appointed data stewards, effectively incorporates a sense of responsibility making data quality a part of the day-to-day tasks.

Problems with ambiguity in data were most evident with regards to the number of redundant fields contained within some of the master data objects. The lack of documentations and proper management of metadata, one respondent said, made it hard, in case of a change request to extend the data model with an additional field, to verify whether this attribute already exists. It is argued that this is due to the lack of standardization which in turn relates to the previously mentioned challenge of reaching a *common data definition*. Nonetheless, Vilminko-Heikkinen & Riikka (2013) stresses the importance data standards as this enabled organizations to better adjust to changing business requirements.

Who will then oversee the management of data quality? There is consensus in literature that this is the responsibility of the data governance organization. Ballard et al (2013) in turn argues that for proper management of the DQ it is the responsibility of the master data governance council to issue data quality policies and metrics. Where typically appointed *data stewards* will be responsible for implementing these metrics into the organization as well as following up on them (Radcliffe, 2007).

In this context Consultant 2 emphasizes the importance of understanding what business requirements are placed on data, in order to understand the prioritization among the data quality dimensions. From interviews it became clear that the desire for an increased reporting capability had required them to address the deficiencies in data. With this in mind, Consultant 2 argues it is important that data is cleaned so that technical keys can be deduced which becomes important in case enterprises want the ability to follow up on customer on a global basis.

With master data in NCC being spread across multiple heterogeneous data sources this in turn has resulted in an array of definitions for the same master data. There are different techniques of addressing this problem, as suggested by both Consultant 1 and Loshin (2011), is the use of a *dictionary* where a business term (or master data) is listed in one column and its various aliases in another along with a column explaining its application area. Unlike the previously failed attempt of having a centralized dictionary, this one will focus on establishing a dictionary with a specific focus on master data. It is also important that there are appropriate processes in place for ensuring that information is always up to date.

The role of stewards for improving data quality is briefly explained in the literature review; they are the ones that carry out and ensure that proper management of data is made. The relation between data governance and data quality has been elaborated on in the literature study and shows that DG and DQ have a symbiotic relationship. Where it is the task of the data governance organization to ensure that quality is on an acceptable level and that master data related policies are enforced in the organization. However, as most DG initiatives are facing the challenge of successfully following through (Loshin, 2009), having formulated policies for master data will be of little good if there is not a governance organization in place to enforce them and make sure that they are adopted by the rest of the organization (Plotkin, 2014).

“If you do not measure it, you cannot manage it” (Consultant 1). Since one of the most common reasons for engaging in a MDM program is to improve the data quality (The Information Difference, 2010) knowing what problems exist and keeping track of the progress in solving them is vital. Without agreed upon metrics for monitoring and quantifying the positive impact of having a more data quality conscious and actively working with it, Radcliffe (2007) argues that the benefits of the MDM program may become eclipsed by a cost focus in case it fails to demonstrate added value to stakeholders. Since implementing MDM is a costly journey, however with a huge upside, the challenge as expressed during interviews is to obtain executive buy-in and retaining this through the years as it is important to make stakeholders aware that it is a multi-year program and one which in the long run may provide a competitive edge. This, as well as for many other organizations, is a challenge facing NCC and with construction companies in general not being at the forefront, it is imperative to communicate the benefits of MDM in a way that makes it understandable for all stakeholders.

As stated before, metrics serve not only to provide an objective view of the current state of data quality (Radcliffe, 2007) but to also help ensure that the vision with the MDM initiative and its alignment with strategic objectives and so that they are measurable (Consultant 2; Consultant 3). It is important that the initiative is anchored in what the organization is trying to achieve. For example, if the aim is to come closer to the user then taking this goal and deriving from it requirements that must be fulfilled. In the same manner, from these requirements the metrics are then derived (an example of this is depicted in Figure 18)

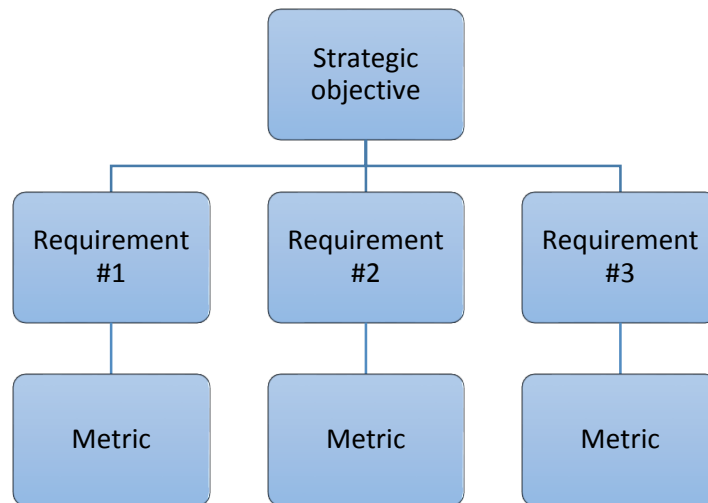


Figure 19: An example of deriving metrics from strategic objective

6.2.2 Common data definition

During interviews with NCC employees one of the most apparent challenges were the ability to obtain consensus regarding the data model. Years of operating in silos with each business areas having had their own business processes and repositories have yielded a range of definitions for some of the master data entities e.g. customer. This phenomena is highlighted by Fisher (2007) who argue that in order to improve one’s data it is important to agree on the definitions of master data. Also agreed by Wolter & Haselden (2006) who explaining that it is important for enterprises contemplating to embark on MDM to understand what makes up master data, as all fields contained by a master data entity need or should not be mastered, a view also shared in (Vilminko-Heikkinen & Pekkola, 2013).

In essence, the lack of coordination amongst stakeholders, inability to resolve disputes relating to data and the inadequate degree of senior executive involvement during these initiatives intersects with what literature attributes to insufficient governance. Where, as an initial step, establishing governance council (or “Center of excellence”) consisting of both IT and business data owners having the final say in master data related conflicts (Fisher, 2007; Wailgum, 2007; Cheong & Chang, 2007; Patton & Wentz, 2010). Furthermore, the imperative of having senior executive sponsorship involved with the governance is important as these empowers the governance initiative to act on data as engaging in governance in NCC will entail a cultural change which likely will face resistance. This view is shared by Consultant 1 who emphasizes the importance of giving the governance initiative “teeth” through decision mandate and needed resources.

Furthermore, if we recall from the empirical study problems stemming from having several stakeholder present with varying opinions in how data should be created difficulties. In this regard insufficient scoping and lack of appointed roles and responsibilities is thought to cause this. In terms of scoping, as previously argued by Consultant 1 tying the initiative to a business problem and taking a phased approach. Moreover, the importance of the council is evident as problems due to lacking direction and coordination, especially having top management articulating requirements, has been an additional source for failed initiatives.

The use of an industrial data model for some of the master data entities composed by some of the most common attributes used in a particular area (in this case construction), is considered by Consultant 1 as an enabler for reaching consensus. Remarks and observations, leads one to believe that discussions are halted due to irreconcilable differences, while Consultant 1 instead argues that most of the time this is not the case. Through the use of a standardized model they

select the fields they intend to master and thereby overcoming any hurdles caused by year of operating in silos resulting in varying business lingo.

6.2.3 Successful change initiative

“Those who do not remember the past are condemned to repeat it” – George Santayana

The importance of understanding why previous attempts to address the master data has failed will prove to be a valuable source of reference when embarking on future activities related to the management of data, to avoid that mistakes from the past are repeated.

From Figure 13 a number of problem variables were identified, for which some of them suggested solution were given in accordance to best-practice. One of the perhaps more notable problem variables was that of commitment. This of course is a complex matter, with many factors having either a positive or negative influence to it. From our analysis of the empirical data in conjunction with the literature review a number of variables found to impact the commitment were recognized such as: lacking vision, hardness of demonstrating business value and misplaced focus due to lacking involvement from business.

A pre-requisite for success with MDM is the existence of either a business problem or a vision, which in turn requires the need for an MDM vision (Radcliffe, 2007). This is agreed by both Consultant 1 and Consultant 2, who argues that without tying yourself to a broader initiative benefiting from MDM, the efforts are deemed to get nowhere. With previous initiative struggling to establish a sense of direction and vision for the area, this further strengthens the belief that previous initiatives had been initiated before understanding which business problem that is driving the need for improved data.

Clearly communicating and educating the organization for the benefits of having proper data management is important in order for stakeholders on all levels of the enterprise to appreciate and understand the importance of good data quality. Having a communication capability is something one would expect to be provided by the data governance organization (Consultant 1).

Years of business lingo, Consultant 1 argue results in business areas having a tendency to consider themselves complete different from the other business areas, while this is many times not the case. And so acknowledging this difference of definitions highlights the need for 1) standardization of data definitions and 2) talking to the business areas in their own language.

As previous attempts of addressing the master data have consisted mostly of IT people, this suggests that there is a need to, in a higher degree, involve the business. In this regard the literature speaks of establishing a governance council where policies, ownership and accountability for the initiative are decided (Patton & Wentz, 2010). Ideally the executive sponsor should chair this council to provide the council with sufficient authority which is important, especially for MDM as it entails a cultural shift for many companies which is often not easily accepted. The role of the business executive will be to provide a business point of view in order to assure the initiative is aligned with the overall business objective. Of course there is only so much time an executive sponsor have allotted to the council; the role of the executive sponsor will be to evangelize the need for MDM to support the business vision and communicating the need form in this case, MDM. However, a piece of warning is given by Dyché & Nevala (2008) to not expect too much from the sponsor when it comes to the practical implementations of the governance organization and other details relating to the MDM initiative. However, against the back of the business vision, a set of policies, metrics and principles must be established to in order to assure vision and to report back to the business leaders the progress (Consultant 2; Cervo & Allen, 2011).

Furthermore, working with change management will be another critical component to MDM. Although there is a generic change management processes promoted by the PMO this need to be enforced and followed up through appropriate processes to address the changes on all levels (Radcliffe, 2011). Consultant 2 agrees with the importance of change management for MDM, and further argues that this must be the responsibility of the project steering group. The need for having a governance council establishing a chart of roles and responsibilities in a matrix form

e.g. RACI is considered a key factor considering a number of remarks obtained during interviews and from observations. With no consequences for those not deciding to adhere to agree on changes. Consultant 1 and Radcliffe (2011) in this regards of the use of levers (also known as carrot and stick) to encourage a certain behaviour. Having the executive sponsor provide resources such as mandate to influence the interaction with data, or a budget to empower individuals to stick through with the interactive is crucial (Consultant 1).

6.2.4 Challenges and success factors

The discussion above can be condensed into a number of key challenges, although not exhaustive, that can help NCC assure a successful MDM program.

Master data governance. There is consensus among MDM practitioners that you cannot have an effective MDM without governance. This is evident from NCC's master data hub GR, which based on the interviews suffers from a number of quality deficiencies due to factors such as: lack of standardization, ambiguous ownership structure of data, to name a few. The lack of appointed data owners for some of the master entities at NCC has created confusion whenever data anomalies are encountered as it is not clear who is accountable to address it. During the course of the MDM program hundreds of decisions related to defining and changing master data are made (Wolter & Haselden, 2006). This requires a well-defined decision-making body the MDM initiative can crumble due to political interest hindering effective decision making.

In addition to establishing a formalized governance organization, another mission for the governance board will be to establish a set of master data principles and policies, anchored in business strategy, to ensure that master data is managed in such a fashion that it will help the enterprise in reaching its business objectives. For example one such principle could be to establish the role of data as an asset, and as such, should be treated as any other corporate asset.

Data stewardship. Establishing a master data governance is one of many crucial steps to ensure a successful MDM program execution. However, as described from the literature study, without data stewards to enforce principles and policies within the operating organization master data related activities, having well-defined data policies will provide little value; this is also one of the major reasons why data governance have historically failed to follow-through.

Furthermore, based on the results and the above discussion the role of data stewards entail deucing, from established data policies and principles, a set of procedures, business rules and data quality metrics that needs to be tracked by the data steward. In NCC's case appointing accountability for master entities becomes a key activity, as many of the shortcomings has persisted due to the absence of accountability. Friedman (2007) stresses that ownership must not be confused with stewardship, where ownership of a master entity implies control for the data. For instance, there could be a data steward appointed to the charts of accounts, however the ownership would likely lie with e.g. CFO.

Ensure program commitment. There are often a range of different stakeholders involved or affected by a MDM program and each perhaps with their own objective. The importance of senior management buy-in is perhaps most obvious as they are needed in any enterprise-wide initiative and will fill an important role as MDM champions communicating the benefits of MDM to the rest of the enterprise. Loshin (2011) argues that it especially in the beginning of a MDM program is important to engage business managers from various LOB to demonstrate how MDM can help them make their day-to-day tasks effectively. Seen in Figure 13, the importance of senior management commitment has indeed been a concern at NCC which further is a prerequisite for conducting any enterprise-wide initiatives (Loshin, 2009). Meanwhile, it is the responsibility of master data champions to communicate the benefits it can provide.

Establish metrics and continuous value display. The motivation for defining metrics are two-fold – firstly they help provide an objective point of view for the current state of data quality.

Awareness for the current state of master data quality remains varies between employees which both indicates insufficient communication capabilities as well as a confusion for what good and bad quality in data looks like. Secondly, metrics enables managers to trace the progress in data and e.g., customer satisfaction, which is the basis on which senior management decisions whether to continue or terminate the program is taken.

The challenge of many MDM initiative is the ability of the organization to continuously demonstrate its value contribution to sponsors and key stakeholders. And with MDM not being a one-time effort the cost focus may eclipse the benefits of structured management of master data. Based on previous cases and the observations made at NCC one of the contributing factors of unsuccessful master data initiatives has been to prioritize the efforts, and thereby focusing on high-impact business area to build out from, rather than having a “all or nothing” approach. Both Consultant 3 and Consultant 1 in this regard emphasis the importance of focusing the efforts on a smaller and more manageable scope of great interest from the business and using that as a gateway to lift the awareness of the benefits of MDM.

Business ownership for master data. Despite the many benefits coming from a MDM environment Consultant 2 argues that in fact only a small number of organization follow-through. As most MDM initiatives tend to be IT driven Consultant 2 sees the risk of being able to get the business onboard a MDM imitative if, as also argued by Consultant 1, it is not clearly tied to an business problem by which trusted and consolidated master data can aid in the business problem or opportunity. From the results it can be seen that some of these shortcomings resides within NCC where activities requiring business expertise instead are executed by IT. In addition, previous attempts to address master data were unsuccessful due to it being to IT-centric with discussion converging into respective participant’s technical expertise domain.

Technology neutrality. Consultant 2 argues that one of the biggest problem and reason or failed MDM initiatives have been in those cases where a particular product has been pre-determined to be used as a master data repository, leading to discussions focusing on the product and less on the soft factors such as governance and mapping the information need without being limited to the structure of a particular vendor. Tuck (2008) takes this one step further emphasizing that MDM is not an IT exercise and that the most important factor affecting success of an MDM initiative is the implementation of appropriate data governance and stewardship processes and less on the technical architecture. Remarks, addressing this, were made during interviews essentially pointing to the fact that challenges lies with organizational change than technical architecture.

Change perseverance. Standardization of master data definition and processes altering the data, is likely to introduce big changes for most organizations. In the context of NCC this becomes especially true, with business areas having had a high degree of agility and freedom in regards to its data management, stemming from the fact that most of NCCs business areas operating as vertical organizations. Therefore the MDM initiative may experience resistance as they may not agree with the changes proposed. A common pitfall in this regard is to allow for exceptions by letting certain business units to have their own local setups for them to accept the MDM approach; as was the case at Sandvik, eventually making them unable to follow-through with its MDM program.

The imperative of communication is evident as seen from the literature study; Kotter (2007) emphasizes the criticalness of communicating out the vision with any change initiative and making it clear how not making the change will lead the enterprise to not being able to fulfill its corporate objectives. In the context of NCC, this could for instance mean that they fail to meet the strategic objective to better understand its customers which in turn makes them unable to come closer to their customers.

Communication capabilities. One of the challenges inherent in the NCC organization is the lack of communication across LOB which historically have acted as isolated information silos. This in turn have hindered NCC ability for knowledge exchange, as evident from the number of systems with overlapping data. One respondent gave the author an example of a, considered transaction-heavy system, where data created at system A where needed by a system B, with both system A and B storing locally storing the same data. The respondent had in this case flagged for this redundancy with no success. Furthermore, by improving the communication, LOBs can make use of existing services structured within e.g. a service catalogue rather than having a palette of different solution for achieving the same task, which could have negative impact on data quality. Consultant 1, agrees with the importance of communication and argues that this is perhaps one of the biggest challenges of an MDM program; therefore as part of the data governance mission is to provide communications capabilities to both retain stakeholder commitment and educate the organization of the benefits of managing their master data.

6.2.5 Tying the knot

In literature countless references to data governance as crucial for MDM success and/or also referred to as the “glue” keeping the different dimensions together. Permeating the challenges identified and discussions made is the notion of data governance, as most challenges either are attributed to the lack of appointed roles, articulated direction or accountability. Given the nature of governance being a control discipline many of the shortcomings have emerges due to the lack of a more formal approach to how data is managed and instilling the view that MDM is not a one-time effort but rather one that is continuous and followed up one. In fact, according to a survey made by The Information Difference (2010) it revealed that 88% of those planning to implement MDM considered implementing DG either before or together with MDM a success factor.

Once the case for the data governance organization has been accepted an equally important question then is – who are the roles involved? Cervo & Allen (2011) insist that the typically found roles in the governance organization need not necessarily be held by individuals of certain titles, on the contrary, focusing on identifying the “doers” in the organization. Neither should it be assumed that these roles appointed are full-time positions, according to (Loshin, 2009; Berson & Dubov, 2011, Plotkon, 2014).

A number of references to different roles have been made during the course of the discussion which are depicted in Figure 20. This is high-level structure based on the description provided in literature. In Appendix A.1 descriptions of the different roles are given.

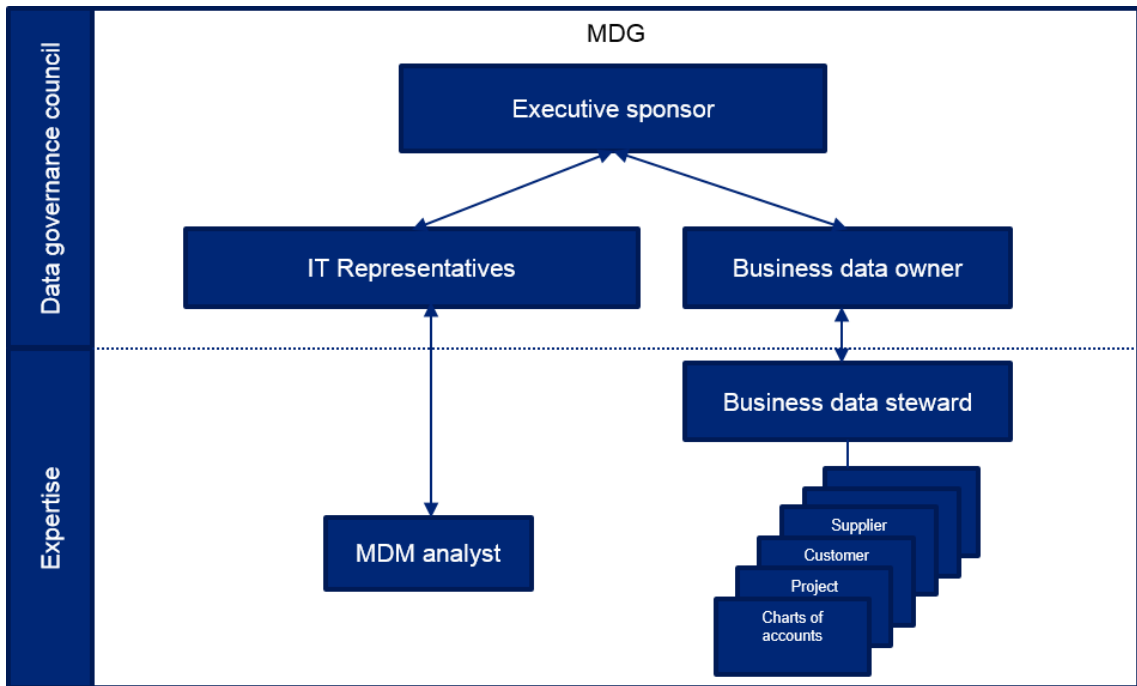


Figure 20 Governance structure derived from (Cervo & Allen, 2011; Loshin, 2009; Dreibelbis et al, 2008)

7 Conclusion

During this section a summary of the results found throughout the course of the thesis is presented, as well as the recommendations for NCC going forward

7.1 Recommendations

First research question:

- “What shortcomings and challenges are there in the way master data is managed at NCC today and what causes them?”

It is evident that NCC are facing a number of challenges, both technically and organizationally in order to successfully implement and realize the benefits of MDM. However, it clear that the toughest challenges facing them are tied to the governance and organizational challenges.

Notably, one of the apparent problems with MDM is the range of definitions making it a tough topic to fully understand. To a great degree this is also the case at NCC, where among NCC respondents seem to be awareness of the difficulties created by the straggling master data. The knowledge of MDM is as expected fairly low which suggests that there is a need to educate the organization of the concept, on a more detailed level anchored in the organization.

From analyzing the data acquired from both interviews and literature study three major challenges (or themes) are derived:

- Data quality
- Common data definition
- Successful change initiative

Second research question:

- “With regards to the identified shortcomings and challenges which are the success factors enabling NCC to improve their management of master data?”

As stated in §1.3 Delimitations, focus of this thesis has been primary on the governance and organizational dimensions of MDM. Where failing to properly manage these dimension according to literature is one the most common causes for MDM programs not being able to follow-through. This has been the case for NCC where previous attempts embarking on MDM have not been able to successfully follow-through, in part, due to inability to generate a compelling business case and agree on a target data model.

Together with the constructed graphs and literature of the fields following success factors are derived:

- Master data governance
- Data stewardship
- Ensure program commitment
- Technology neutrality
- Establish metrics and continues value display
- Business ownership
- Change perseverance

From the challenges and proposed success factors we can also conclude that a majority of the challenges are related to the absence of a data governance function focusing on master data as well as appointed roles responsible for the different master data entities. For this reason a reference model for the governance organization is proposed (see Figure 20) encompassing following roles:

- **MDG council** – Regarded as “center of excellence” consisting of both business and IT managers preferably chaired by the senior sponsor. Their primary responsibility will be

to address issues in master data that are escalated to them by data stewards. They will also be responsible to track the progress of the initiative and to ensure that expectations from the business are met. Having a clear understanding of how, to the business, success is measured is vital as NCC currently do not have in place any metrics (or KPIs) in place which can cause the initiative to crumble if it fails to demonstrate improvements in areas concerning the business.

- **IT coordinator** – It is important that appropriate expertise is present during council meetings to enable them to make informative decisions. In this case the IT-coordinator ensures appropriate expertise from IT are present to give input on solution design and to highlight potential limitations seen from the technical side.
- **Business data owners** – Perhaps one of the biggest challenges of NCC will be to institute ownership for the different master entities e.g. customer, project, supplier etc. Although informally there may be owners of data this must be explicit. How the ownership structure for the various entities will be designed is not easily solved, partly due to the two dimensions: business area (i.e. NCC Construction, NCC Housing, NCC Property development, etc.) and geography (i.e. NCC Sweden, NCC Norway, NCC Denmark, etc.). How this will take form is outside the scope of this thesis but will be crucial success of MDM. Their responsibilities entails, among others, to raise any issues escalated to them by business data stewards and to delegate work down to business stewards which are then followed-up and presented during council meetings.
- **Business data stewards** – rather than solving issues with data whenever they occur NCC, and other alike, should instead focus on proactively prevent poor quality data to be inserted in the first place. For this reason deriving master data related policies and enforcing them by means of data controls at the point of creation. Hence, the responsibility of the business data stewards with support of the operational data stewards is to ensure that standards and policies are enforced and accordingly adhered to.
- **MDM analyst** – having an individual with expertise in master data responsible for the data quality administration while maintaining an oversight of data and advising on opportunities of data harmonization and standardization, is especially important in the case of NCC. As the field of MDM is steadily maturing, hence, literature remains scarce on how to begin. The MDM analyst will as such both support the MDG organization with input based on best-practice and data insight while also assist, on a project-by-project basis, project teams to ensure that awareness of the master data concept is kept and help assess impacts of a change or solution.

7.2 Reflection

In this thesis we set out to understand why the master data hub at NCC referred to as GR had failed in regards of becoming trustworthy authoritative source of master data. The significance of GR to NCC is evident seen to the share amount of systems consuming its data. Initial interviews with NCC employees with support from literature suggested that although there are technical challenges with respect to its inherently complex IT landscape. It appears the awareness for these challenges were somewhat better understood, in comparison to the challenges residing with people, data quality and governance.

Using causal opinion graphs to depict the relations among the identified problem variables proved most challenging due to large variances in opinions and where finding a common denominator (i.e. underlying cause) were hard to derive and subsequently required more time than initially anticipated.

In addition holding interviews at times proved also to be unexpectedly hard depending from respondent to respondent and if more time were provided having a second iteration of interviews could have further strengthen the reliability or/and given further insight to the problems.

Regardless, through this thesis a number of areas of interest have been highlighted and I am confident that the challenging areas highlighted will be useful for NCC to consider if they decide to embark on MDM going forward. It is my strong belief that NCC would strongly benefits from better management of their data, not only for operational efficiency as this decreases the non-value adding activities that currently must be executed due to the poor quality in data, but also provides chances of cost reductions through increased understanding of inventory and material costs and thereby eliminating potential waste.

7.3 Future work

Although this thesis have raised the awareness for some of the challenges facing NCC in its work to establish MDM, this work can further developed/extended. Suggestions for future work entails:

- Increasing the number of respondents. Since opinions were are only gathered from the Swedish office a possible continuation of this thesis would entail increasing the number of respondents and the geography from which these respondents are selected i.e. conduct interviews with respondents across all NCC subsidiary offices in the Nordic region. Doing this is likely to increase the reliability and generalization of the results.
- Another interesting future research work, continuing on this thesis would be to focus on one of the master data entities, for example, *customer* or *project* and assess the current state of this entity and processes relating to it. The purpose of this, in part, is to raise an awareness for some of the shortcomings and consequences provide NCC with a recommendation for a target state and the actions required to realize it.
- Extending methodology used in this thesis by complementing with a case-study performed at a company similar to NCC with an already established MDM practice. In doing this one can get a deeper insight to how MDM is practically implemented in an organization, and how some of the identified challenges at NCC have been addressed in this case-study company. As literature remains scarce of practical implementation details.

8 References

- Ballard, C. et al., 2013. *An IBM Redbooks publication*. [Online]
Available at: <http://www.redbooks.ibm.com/abstracts/sg248059.html?Open>
[Accessed 14 April 2014].
- Ballard, C. o.a., 2013. Master data governance and stewardship. i: *Aligning MDM and BPM for Master Data Governance, Stewardship, and Enterprise Processes*. USA: International Business Machines Corporation, pp. 15-26.
- Berson, A. & Dubov, L., 2011. *Master Data Management and Data Governance*. 1st red. USA: The McGraw-Hill Companies.
- Butler, D., 2011. *Master Data Management: An Oracle White Paper*, Redwood Shores, CA, USA: Oracle.
- Cervo, D. & Allen, M., 2011. *Master Data Management in Practice: Achieving True Customer MDM*, Hoboken: John Wiley & Sons, Inc..
- Collis, J. & Hussey, R., 2003. *Business research: a practical guide for undergraduate and postgraduate students*. 2nd red. u.o.:Palgrave Macmillan.
- Das, T. k. & Mishra, M. R., 2011. A Study on Challenges and Oppportunities in Master Data Management. *International Journal of Database Management Systems (IJDMS)*, May.3(2).
- Data Governance Institue, u.d. *About Data Governance*. [Online]
Available at: http://www.datagovernance.com/adg_data_governance_definition.html
[Accessed 6 4 2014].
- Dreibelbis, A., Milman, I., Run, P. v. & Hechler, E., 2008. *Enterprise Master Data Management: An SOA Approach to Managing Core Information*. s.l.:IBM Press.
- Dubov, L., 2014. *IT Strategy driven MDM (part three)*. [Online]
Available at: <http://www.ibmbigdatahub.com/blog/it-strategy-driven-mdm-part-three>
[Accessed 19 July 2014].
- Dyché, J. & Nevala, K., 2008. *Ten Mistakes to Avoid When Launching Your Data Governance Program*. [Online]
Available at: http://www.sas.com/resources/whitepaper/wp_63774.pdf
[Accessed 21 April 2014].
- Eckerson, W. W., 2002. *Data quality and the bottom line*, u.o.: 101communications LLC.
- Fisher, T., 2007. *Demystifying Master Data Management*. [Online]
Available at: <http://www.cio.com/article/2439152/data-management/demystifying-master-data-management.html>
[Accessed 14 10 2014].
- Flick, U., 2009. *An introduction to qualitative research*. 4th red. London: SAGE.
- Friedman, T., 2007. *Best Practices for Data Stewardship*, u.o.: Gartner.
- Haug, A., Zachariassen, F. & Liempd, D. v., 2011. The costs of poor data quality. *Journal of Industrial Engineering and Management*, 4(2), pp. 168-193.
- IBM, 2007. *IBM Master Data Management: Effective data governance*. [Online]
Available at: <ftp://public.dhe.ibm.com/software/uk/itsolutions/information-management/information-transformation/master-data-management/master-data-management-governance.pdf>
[Accessed 14 4 2014].
- Kernochan, W., 2006. *Master Data Integration: A Key Enabler of Master Data Management*, Lexington, MA: Infostructure Associates.
- Khatri, V. & Brown, C. V., 2010. Designing Data Governance. *Communications of the ACM - Amir Pnueli: Ahead of His Time*, January, 53(1), pp. 148-152.
- Kotter, J. P., 1996. *Leading change*, u.o.: Harvard Business Press.

-
- Kotter, J. P. & Schlesinger, L. A., 1979. Choosing Strategies for Change. *Harvard business review*, 57(2), pp. 106-114.
- Ladley, J., 2012. *Data Governance*. 1st red. Waltham, MA: Morgan Kaufmann.
- Le Duc, M., 2007. *Metodhandbok som tankekarta*. [Online]
Available at: <http://www.leduc.se/metod/Induktion,deduktionochabduktion.html>
[Accessed 12 May 2014].
- Loshin, D., 2009. *Master Data Management*. s.l.:Morgan Kaufman.
- Marsh, R., 2005. Drowning in dirty data? It's time to sink or swim: A four-stage methodology for total data quality management. *Journal of Database Marketing & Customer Strategy Management*, Volym 12, pp. 105-112.
- Maxwell, J. A., 2004. Using qualitative methods for causal explanation. *Field methods*, 16(3), pp. 243-264.
- Messerschmidt, M. & Stüben, J., 2011. *A global study on master data management*, u.o.: PricewaterhouseCoopers AG.
- Milanov, G. & Njegus, A., 2012. Using Scrum in Master Data Management Development Projects. *Journal of Emerging Trends in Computing and Information Sciences*, 3(11), pp. 1446-1452.
- Miles, M. B. & Huberman, A. M., 1994. *An Expanded Sourcebook*. 2nd red. u.o.:SAGE Publications.
- Moseley, M., 2008. *Information management*. [Online]
Available at: http://www.information-management.com/specialreports/2008_105/10001919-1.html
[Accessed 25 7 2014].
- OECD, 2004. *OECD Principles Of Corporate Governance*, Paris, France: OECD Publications Service.
- Olson, J., 2002. *Data Quality: The Accuracy Dimension*. 1st red. u.o.:Morgan Kaufmann.
- Otto, B., 2011. Data Governance. *Business & Information Systes Engineering*, August, 3(4), pp. 241-244.
- Patton, D. & Wentz, R., 2010. *Publications: Master Data Governance, Model and Mechanism*. [Online]
Available at: <http://www.pwc.com/us/en/increasing-it-effectiveness/assets/mdm-model-and-mechanism.pdf>
[Accessed 15 4 2014].
- Pipino, L. L., Lee, Y. W. & Wang, R. Y., 2002. Data Quality Assessment. *Communications Of The ACM*, 45(4), pp. 211-218.
- Plotkin, D., 2014. Data Stewardship and Data Governance: How They Fit Together, In Data Stewardship. i: D. Plotkin, red. *Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance*. Boston: Morgan Kaufmann, pp. 1-21.
- Radcliffe, J., 2007. *The Seven Building Blocks of MDM: A framework for Success*. [Online]
Available at:
<http://www.gartnerinfo.com/mdm2/TheSevenBuildingBlocksofMDMAFrameworkforSuccess.pdf>
[Accessed 17 March 2014].
- Radcliffe, J., 2011. *You need Change management to Succeed With Master Data Management*. [Online]
Available at: <https://www.gartner.com/doc/1289731/need-change-management-succeed-master>
[Accessed 10 August 2014].
- Redman, T. C., 1998. The Impact of Poor Data Quality on the Typical Enterprise: Poor Data Quality has far-reaching effects and consequences. *Communications Of The ACM*, 41(2), pp. 79-82.

-
- Ross, J. W., Weill, P. & Robertson, D. C., 2006. *Enterprise Architecture as Strategy: Creating a Foundation for Business Execution*. Boston, Massachusetts: Harvard Business School Publishing.
- Sammon, D., Nagle, T. & Carlsson, S., 2012. Making sense of the Master Data Management (MDM) concept: old wine in new bottles or new wine in old bottles?. *Journal of Decision Systems*, 21(3), pp. 245-258.
- Sarsfield, S., 2009. *The data governance imperative: a business strategy for corporate data*. Cambs, GBR: IT Governance Publishing.
- Scannapieco, M., Missier, P. & Batini, C., 2005. Data Quality at a Glance. *Datenbank-Spektrum*, Issue 14, pp. 6-14.
- Strong, D. M., Lee, Y. W. & Wang, R. Y., 1997. Data Quality In Context. *Communications Of The ACM*, 40(5), pp. 103-110.
- Svennevig, J., 2001. Abduction as a methodological approach to the study of spoken interaction. *Norskrift*, Volym 103, pp. 1-22.
- The Information Difference, 2010. *How Data Governance Links Master Data Management and Data Quality: An Information Difference Research Study*, u.o.: The Information Difference Company Ltd.
- The MDM Institute, u.d. *The MDM Institute: Independent. Authoritative. Relevant.* [Online] Available at: <http://0046c64.netsolhost.com/whatIsDataGovernance.html> [Accessed 12 4 2014].
- Tuck, S., 2008. Opinion piece: Is MDM the route to the Holy Grail?. *Journal of Database Marketing & Customer Strategy Management*, Issue 15, pp. 218-220.
- Wand, Y. & Wang, R. Y., 1996. Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, 39(11), pp. 86-95.
- Weber, K., Otto, B. & Österle, H., 2009. One Size Does Not Fit All - A Contingency Approach to Data Governance. *ACM Journal of Data and Information Quality*, June, 1(1), p. 27.
- Weill, P. & Ross, J. W., 2004. *IT Governance: How Top Performers Manage IT Descision Rights for Superior Results*. Boston: Harvard Business School Press.
- White, A., 2010. *MDM 'Primer': How to Define Master Data and Related Data in Your Organization*. [Online] Available at: <https://www.gartner.com/doc/1438416/mdm-primer-define-master-data> [Accessed 30 July 2014].
- White, A., Newman, D., Logan, D. & Radcliffe, J., 2006. *Mastering Master Data Management*, u.o.: Gartner.
- White, C., 2007. *Using Master Data in Business Intelligence*, u.o.: BI Research.
- Wiederstam-Paul, F. & Eriksson, L. T., 1991. *Att utreda, forska och rapportera*. u.o.:Liber.
- Vilminko-Heikkinen, R. & Pekkola, S., 2013. *Establishing an Organization's Master Data Management Function: A Stepwise Approach*. u.o., IEEE, pp. 4719-4728.
- Wolter, R. & Haselden, K., 2006. *The What, Why, and How of Master Data Management*. [Online] Available at: <http://msdn.microsoft.com/en-us/library/bb190163.aspx> [Accessed 27 March 2014].

Appendix

A.1 Role description

Role	Description
Master data governance council	
Executive sponsor	Participates in the MDGC meeting and is the driving force and ultimately responsible for the success of the initiative. Furthermore, the sponsor is responsible for allocating necessary funding and resources as well as bringing in a cooperate perspective to ensure the initiative is aligned with the business strategy.
Business data owner	Depending on scale, this individual is responsible for one or multiple master data entities and reviews/escalates issues related to data entities to the council. It is the responsibility of the business data owner to delegate work packages to data stewards. Changes to data entities is communicated and coordinated with the IT coordinator
Subject expertise	
MDM analyst	The role of the MDM analyst is twofold: first, supporting the governance function by providing a best practice point of view as well as data insights. Second, as part of the MDM program assisting in the projects to ensure they align to data quality and architectural requirements, and to help assess impact of changes to master data.
Business data steward	Based on the directives provided by the business data owners, they are in in charge of ensuring that master data standards are defined and adheres to policies and quality requirements.
Operational data stewards	Individuals responsible for maintaining the master data, e.g. making sure master data standards are followed. Preferably they work closely with the business data steward, using information such as, #issues and # changes requests to form recommendation on areas in need of addressing.

A.2 Deloitte's master data identification framework

Criteria	Description	Rating
Shared	Is the data used by more than one business process/system?	0 – Data used in a single system/process 1 – Data used by two systems/processes 2 – Data used by more than two systems/processes
Value	The element is fundamental to a business process, subject area, or business system.	0 – Data is useful to individuals only 1 – Data is critical to a single business process 2 – Data is critical to multiple business processes
Volatility	Data modification behavior	0 – Transaction data 1 – Reference data 2 – Data added to or modified frequently, but the data is not transaction data
Total		

Results

0-2	Attribute is not master data (or any criteria is rated 0)
3-4	If any criteria is rated 0, attribute is not considered master data. Otherwise, attribute minimally meets criteria for master data and further investigation should be considered
>4	Attribute is master data