An analysis of the benefits and issues in the development of an Enterprise Data Catalogue

By

Xuemei Wang

300085084

Submitted to the School of Information Management,

Victoria Business School,

Victoria University of Wellington

In partial fulfilment of the requirements for the degree of

Master of Information Management

17th October 2014

Declaration

I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.

I have used the APA convention for citation and referencing. Each contribution to, and quotation in, an analysis of the benefits and issues in the development of an Enterprise Data Catalogue, from the work(s) of other people has been attributed, and has been cited and referenced.

This paper is my own work.

I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

I acknowledge that copying someone else's assignment, essay or paper, or part of it, is wrong, and declare that this is my own work.

Signature(s) Xuemei (Mei) Wang Date 15 / October / 2014

Preface

I would like to express my deepest appreciation to all those who provided me the possibility to complete this case study report:

The participants in the case study Ministry, whose time and contributions are invaluable to this case study.

A special thank you to the case supervisor – Tony Hooper, who provided me with support, direction and guidance throughout the case study process.

I would also like to acknowledge with much appreciation the crucial role of Philip Worthington at Victoria University library, who provided me with research support.

Abstract

Purpose – The purpose of this case study is to examine the rationality for taking an enterprise approach towards metadata's development. It outlined and discussed the benefits and challenges involved from both theoretical and practical perspectives.

Methodology – The data for this case study is collected through a one on one interview method. The findings were discussed and analysed with other literature.

Findings – Taking an enterprise wide approach towards the development of metadata offers both short-term and long-term benefits to the Ministry. In particular, the short-term benefits include efficiency in time to answer, re-use and re-purposing of existing information. The long-term benefits are the ability to operate at a strategic level with its data and information.

On the other hand, the different level of understanding towards the concept of metadata at present, in addition to a lack of a clear guardianship from a data perspective has lead to challenges in gaining a level of active contribution and ongoing maintenance.

This report suggested that it is essential for the Ministry to maintain a constant marketing style of communication to promote the understanding of the metadata and to incorporate the principle of data stewardship in managing the component of metadata mapping.

Value – This case study provides value by explaining the concept and outlining the key benefits and issues involved to those organisations discovering the topic of metadata management for the first time. This case study adds additional value by offering rare insights into taking an enterprise approach towards the development of a business metadata, offering practical learning and suggestions for other organisations that are embarking on a similar journey.

Key Words – Metadata, Enterprise Data Catalogue, Information Discovery tool, Data Stewardship, Assets management, Decision-making

Paper Type - Case Study Research

Contents Page

Chapter 1. Introduction

Metadata plays an important role in information management and the decision-making process. As the quantity of information grows, the ability to find desired information in a timely manner has became increasingly important. Metadata supports decision-making processes by facilitating for the fast and efficient retrieval of required information, and assisting the utilisation of existing knowledge and expertise within an organisation. By doing so, it also reinforces the reliability of information management in the long-term.

The purpose of this case study is to examine the benefits and issues in the development of an Enterprise Data Catalogue in a New Zealand public sector organisation (hereafter referred to as "the Ministry").

1.1 Background

The Ministry was established through a merger two years ago. The newly established Ministry covers a broad spectrum of services, ranging from strategic advice on long-term economic growth through to sustainable management of natural resources. It is considered to be a large organisation with over 3000 employees, due to the nature of the services it provides, it also has a very diverse employee base.

The Ministry has committed to a long-term enterprise-wide work plan. The plan aims to bring together a wide range of datasets, that were previously disparate, in a unified and cohesive manner. The Ministry's data and information strategy was developed to support the work plan, and set a path for maximising the contribution of the organisation's data and information assets. The development of an Enterprise Data Catalogue was identified as one of the deliverables in the data and information strategy to support this.

The researcher was brought in to work on the development of the organisation's Enterprise Data Catalogue (EDC), its implementation, and thereafter maintenance. At the time the case study was conducted, the first phase of data collection had just been completed and the EDC is currently being published.

For those organisations still discovering the topic of metadata and metadata management for the first time, this case study provides value by explaining the concept and outlining the key benefits and issues involved.

This case study adds additional value by offering rare insights into taking an enterprise approach towards the development of a business metadata. This is achieved by taking into account the organisational context, when examining the business values for doing so, along with the challenges the Ministry has encountered to date. It also offers practical learning and suggestions for any other organisation that is embarking on a similar journey or is in a similar situation.

Chapter 2. Literature Reviews

Data and information play a vital role in the decision-making process. As the quantity of information grows, the ability to navigate through masses of available data, and locate the right information in a timely manner, has become crucial (Vander Noot, 1998).

Metadata is considered as an essential part of information management. Its purpose is to "support, identify, authenticate, describe, locate and manage resources in a systematic and consistent way" (Technical Committee ISO/TC 46, 2009, p. 3). It thereby facilitates the decision-making process by providing for fast and efficient retrieval of information. By surfacing, then aiding in the utilisation of the existing data and information resources, it also helps to avoid the expensive process of re-collection (Franks & Kunde, 2006).

2.1 Metadata Defined

Metadata is defined as the data that is used for "...describing context, content, and structure of documents, records and their management through time" (Technical Committee ISO/TC 46, 2001, p. 3).

Elements are the foundation of metadata. Depending on the purpose and intended use of metadata, suitable elements may vary. Nevertheless, based on the above definition, Baca (2008) explained the information that metadata usually captures:

- "**Content** relates to what the object contains or is about…
- **Context** indicates the Who, What, Why, Where, and how aspects associated with the object's creation…
- **Structure** relates to the formal set of association within or among individual information objects…" (Baca, 2008, p. 2).

Metadata management is another key concept derived from the definition. There are two interrelated parts to metadata management. Firstly, at a high level, metadata management implies "the implementation of a metadata policy (i.e. principles that form the guiding framework within which metadata exists)" (Westbrooks, 2004, p. 144). This includes coordinating the intellectual activities and required resources, for the creation and manipulation of metadata, such as physical resources and financial commitment (Kurth, Ruddy, & Rupp, 2004). Secondly, from a day-to-day operational point of view, metadata management refers to "a sum of

activities designed to create, preserve, describe, maintain access, and manipulate metadata..."(Westbrooks, 2005, p. 6). The high-level metadata management helps to set the guidelines around the preservation, access and versioning rules. While the day-to-day management of metadata operates within these rules, to ensure the new and existing metadata records stay up to date, meaningful and relevant to the end users. Only together, do they ensure the collection and manipulation of metadata is efficient, consistent and cost-effective.

2.2 Categorising Metadata

"Metadata" is a broad term. It can be used in different situations to carry out different functions. For instance, a web developer might use metadata to refer to information being encoded into HTML meta-tags, which allows easy discovery of a web site. A records archivist might use the term to refer to all the contextual, processing and use of information required, in order to identify the validity and integrity of a record (Gilliland, 2008). A system administrator might refer to metadata as the data that is used to monitor "…data usage, including information on user, security, and access privileges to data and applications" (Shankaranarayanan & Even, 2006, p. 90). The diverse interpretation and use of metadata also leads to different ways of it being categorised. Nonetheless, based on its purpose and use, metadata can be separated into business metadata and technology metadata.

Business metadata is used for "data valuation and interpretation" (Shankaranarayanan & Even, 2006, p. 90). It aims to help business people or nontechnical users, in understanding the data, by providing context to the data in the language of the business users (Inmon, O'Neil, & Fryman, 2008). This type of metadata is often used to carry out the functions of administration, description and preservation of information resources (Gilliland, 2008). Business metadata is especially important for data that is shared by multiple business processes, as it helps to ensure that related information is meaningful to different departments for different reasons within an organization.

Technology metadata is for "system operation and maintenance" (Shankaranarayanan & Even, 2006, p. 90). In a data warehouse, it is used to capture information about the physical characteristics of a database, such as database type

and relationships between tables. It may also be used to summarise the components of the system, such as operating system, networking components and database servers. Technology metadata is often used to describe how a system functions, and the level and type of use of information resources (Gilliland, 2008). Therefore, in a narrow view, this type of metadata is frequently referred to as "a system's data dictionary" (Shankaranarayanan & Even, 2006, p. 90).

2.3 The role of business metadata

Business metadata and technology metadata are inter-related. Firstly, business metadata helps to surface the information that already resides within a data warehouse, by providing business meaning to the data for the data warehouse user or general business user. Then technology metadata provides "technical properties such as table and field structure…" (Hamzah & Sobey, 2012, p. 449) to preserve the quality and consistency of the data itself, ensuring the data is fit to use in a given business context. Secondly, business metadata can be used to document the business justifications for data that is not already stored in a proper data repository i.e. data entered into Microsoft Excel or stored in a Microsoft Access database. By preserving the business context and relevance of the data, it also helps to rationalise the need for it to be data warehoused.

In addition, business metadata also helps to bridge the gap between technology metadata and the use of the data, especially in a data warehouse environment. Technology metadata is a technical necessity in the sense of system operation and maintenance. Conversely, it does not capture the implicit knowledge about the relationship between the data warehouse data and the business process, system or products (Stefanov & List, 2007). As demonstrated in Figure 1, without the business metadata to provide information on the what, who, why and how in a data warehouse environment, the use and context of the data is up to individual interpretation. Without business metadata to bridge the gap between the use of data and its appropriate context, there is the possibility of inconsistent data analysis and findings, and subsequently poor decision-making based on that analysis.

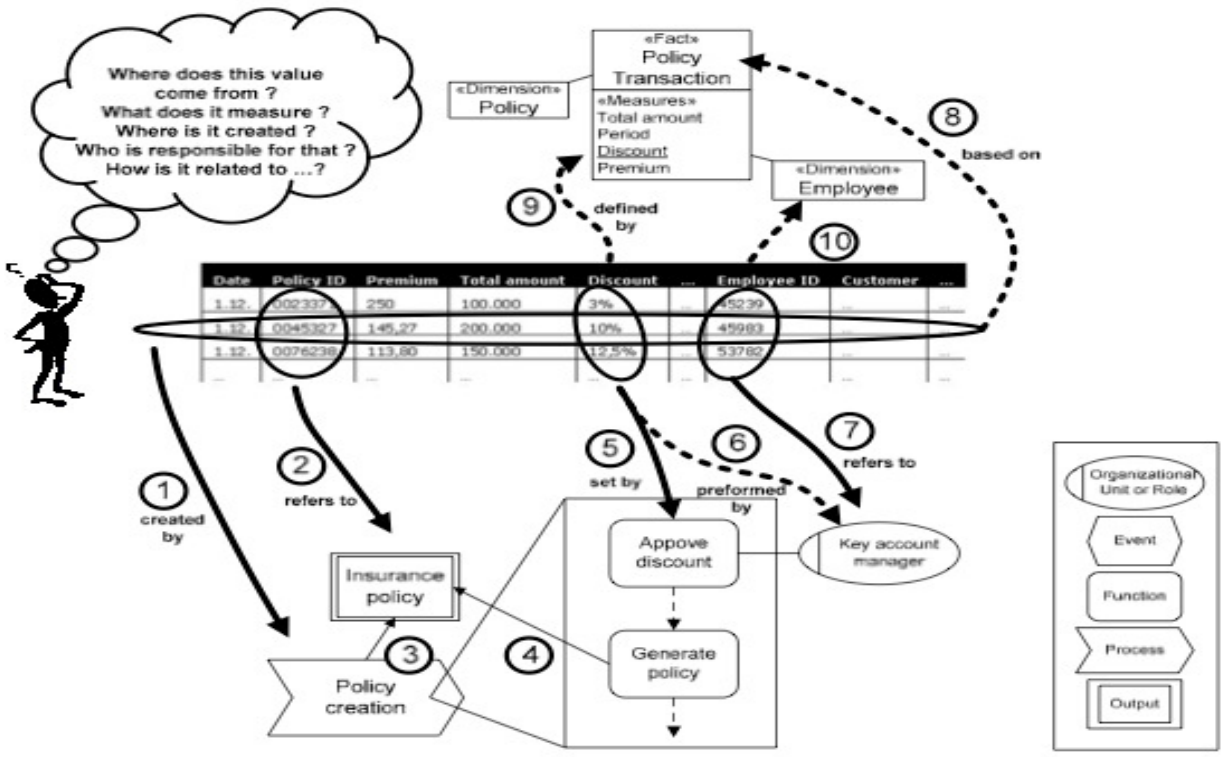

Figure 1 the role of business metadata in a data warehouse (Taken from Stefanov & List, 2007, p. 2070)

Nevertheless, rationally speaking, a carefully designed and well-scoped business metadata, along with a consistent gathering approach would be able to assist in information discovery for both business and technical users at a minimum. Taking an enterprise wide approach towards the designing of business metadata, and collecting the information for an agreed set of metadata elements, would also reinforce the reliability of information management in the long-term.

3. 0 Benefits of an enterprise approach towards metadata design

3.1 Increased accessibility

A consistent metadata can enhance the effectiveness of searching capabilities. Taking an enterprise approach towards the creation of metadata and its collection, means information is tagged and categorised collectively (Roszkiewicz, 2010). This process identifies silos of information that were held by different departments or distributed across several repositories. Surfacing then centralising disparate information, simplifies the process of finding the right people with the right information, and makes it possible for related information to be made available, and accessible to different departments (Roszkiewicz, 2010).

Furthermore, an additional value of taking an Enterprise approach is the ability "to enable data exchange and sharing" (Owens, 2008, p. 372). The process of defining the data used across the organisation, its relationship and local synonyms, means the business activities can be built on a shared knowledge and a common language (Owens, 2008). As a result, it helps to bring an increased clarity to the outputs that are generated by different activities and departments.

3.2 Expanding use of data

Well-structured metadata provides alternative ways for users to get visibility on the available datasets and information. The actual data often has access constraints, for reasons of privilege, security or other barriers. Business metadata can be used as a middle layer between the user and information systems (Shankaranarayanan & Even, 2004). This is possible as metadata provides descriptive information about the data that is stored in a system or repository, without compromising the sensitivity of the data itself. For instance, for sensitive client information, it would record the descriptive detail about the type of information it holds, such as client name, address, contact details, etc., without exposing the actual client data itself. So metadata offers the flexibility for users to assess the suitability of the information against a given business need, prior to the process of requesting access to the data. This leads to better productivity, as less time is spent on searching and validating for the right information, and more time is focused on the actual use of the data and information itself (Roszkiewicz, 2010).

3.3 Supports Master Data Management

An enterprise Metadata standard supports master data management by combining information from different sources in a unified format. Master data is a term used to describe a core set of business-oriented data that can be integrated into an enterprise-wide system or used across multiple business processes (Silvola, Jaaskelainen, Hanna, & Haapasalo, 2011). A typical set of business data could be "parties, including organisation, customer, people, employee, vendors, supplier or trading partners and prospect…" (Silvola et al., 2011, p. 3). Often different teams store a portion of this business data for a particular purpose. For instance, team A records the headquarters address of company X for billing purposes, while team B records a physical site address of company X for logistics purposes, so interrelated information is disjointedly maintained by individual teams incurring separate costs to support it. By tagging and categorising information and data collectively at an enterprise level, it provides a foundation for better understanding of information linkages and building a "single version of truth" (Silvola et al., 2011, p. 3). In addition there is a cost benefit, through promoting sharing, consolidation, and reuse of knowledge and expertise within an organisation.

4.0 The issues of an enterprise approach towards business metadata design

"Metadata is like interest: it accrues over time" (Gilliland, 2008, p. 18). A well-designed metadata with a consistent gathering technique can facilitate data mining, knowledge sharing and improve data accessibility within an organisation. However, the challenges involved in metadata development and management at an enterprise level are far from trivial.

4.1 Diverged Metadata standards

The first challenge when it comes to metadata is, to determine the appropriate metadata schema that needs to be applied, in order to best meet the purpose and intended use of it. Many schema sets have been formalized and are recognised as official standards including:

- Dublin Core metadata initiative (http://dublincore.org), which is one of the most well-known metadata standards. It is made up of a set of fifteen document properties (Wu & Li, 2008), which can be used to describe and search across a wide variety of information resources on the Word Wide Web (Gilliland, 2008).
- Electronic Recordkeeping Metadata standard which sets out a systematic approach to ensure the information managed in business systems and applications has meaning and is application-independent (http://archives.govt.nz);
- ISO15489-1 Information and Documentation – Records management standard (Technical Committee ISO/TC 46, 2001, p. 1) sets a framework "to enable standardized description of records and critical contextual entities for records…"

Depending on the organisation's specific requirements or needs and the type of information object in scope, customisation of a best-fit schema may be required. This was demonstrated in a case report from a digitized historical fashion collection project. Zeng (1999, p. 1193) expected that the cataloguing format for online fashion could follow the principles and standards of object description, which had been established by information professionals. Various standards were examined as part of the case study to determine their suitability for use. However Zeng (1999) found

that the examined standards were primarily designed for document and document-like objects. As a result, in order to meet the organisation's unique needs, supplementing a best-fit schema with additional elements was required. Lubas, Wolfe and Fleischman (2004) also noted and highlighted in their case report for creating metadata practices for an OpencourseWare project, that adding additional metadata elements was required in order to "… accommodate needs that were not addressed in the standard's specifications" (Lubas et al., 2004, p. 138).

As demonstrated above, to evaluate a metadata standard against organisational needs, and then determine when to supplement it with additional metadata elements as well as what type of elements are the best fit, can be a challenge. By extension, in taking an enterprise wide approach towards the metadata standard, the challenge not only lies in identifying the appropriate metadata standards, but also finding the right balance at an enterprise level and across multiple disciplines.

4.2 Management of metadata

Interoperability among existing metadata schemas is another challenge. From a data usage point of view, during the conversion of data to information, and then information to knowledge, there are many interaction points. As a result, multiple business metadata may be captured in different formats, and stored in various repositories for different needs, as discussed by Kurth et al. (2004). In examining the issue, they concluded that one of the fundamental factors of the interoperability issue in metadata management is an inadequately established workflow process. A lack of proper workflow means poor coordination in metadata creation, manipulation, and then in managing output from the metadata processes. This is especially the case when new data is generated due to a new project or system. As a result, inadequate workflow is highly likely to lead to conflicted captions about the data (Kurth et al., 2004), subsequently it may also lead to unnecessary confusion when different users reference different metadata repositories that have different descriptive information about the same set of data.

In response to this issue, Kurth et al. (2004) recommended two solutions. The first was to reach an enterprise wide consensus on metadata mapping decisions. By doing so, stakeholders are engaged early on, which can help to promote sharing, and reuse

of metadata resources later. It is also likely to reduce the costs along with the risks of duplicated effort by different teams. Secondly, Kurth et al. (2004) recommended reinforcing metadata coordination by establishing data stewardship responsibilities in metadata management. A clear accountability for the data helps to govern the maintenance of metadata. And by doing so, it helps to form a point of contact for further enquiries regarding information related to a set of data.

4.3 Metadata Tools

To automate harvesting of metadata poses a technical challenge on its own. Shankaranarayanan and Even (2006) highlighted several emerging issues from a commercial solution point of view, they summarised that:

- "...Little support for business metadata while a strong emphasis on technical metadata…
- Metadata elements tightly coupled with one product or within a suite of products from the same vendor (such as IBM, Microsoft, or Oracle).
- Limited support is offered by way of common Interfaces for metadata exchange, hindering metadata integration across tools…" (Shankaranarayanan and Even, 2006, p. 91).

Nonetheless, in order to determine a suitable technical solution, the following two aspects need to be defined. First of all, "the issues that needs to be addressed in implementing metadata" (Technical Committee ISO/TC 46, 2009, p. 1). By scoping the problem statement for metadata's creation, it outlines the expected benefits, which forms a foundation for evaluating the utility of metadata as well as the information that needs to be captured. Secondly, by understanding the expected benefits, it also assists in carrying out the cost and benefit analysis against a technical solution or the utilisation of any available toolsets.

Chapter 3. Outstanding Questions

The above literature review has demonstrated that, metadata plays an important role in the information management and decision-making process. A carefully designed metadata can simplify the processes of information discovery. It also helps to promote knowledge sharing and reuse of information in an organisation.

Based on the purpose and use of metadata, metadata can be categorised into business metadata and technology metadata. From a theory perspective, the benefits of an enterprise wide approach towards the design of a metadata standard can lead to enhanced search capabilities, expanded use of data, and supports master data management. On the other hand, identifying the appropriate metadata standard, metadata management process, as well as finding a best-fit technology to utilise can be challenges in developing an enterprise wide metadata.

Using the literature review as a starting point, the report has formed the following questions along with their rationale to further examine and validate the benefits and issues in a given organisational context.

Questions	Rationale
1. Explain the main advantages of a corporate approach to metadata's creation, development and maintenance?	By articulating the expected advantages for the data catalogue's implementation, it outlines the current situation and business problems. In other words, this is an advantage, because…?
2. What are the main difficulties one may expect when taking a corporate or enterprise approach in creating, developing and maintenance of business metadata?	By articulating the issues that the organisation has encountered to date, it offers practical insight for the development of the data catalogue.
3. Describe the order you would follow to implement a successful EDC?	Ties back to 5 - The report assumes that the most critical factors would need to be addressed early on. And the item that needs to be addressed early on can also be an indication for the key considerations.

4. How would you measure the success of an EDC project?	Ties back in to question 6. This report assumes that the measure of success would be inline with the expected benefits and thereafter its utilisation.
6. For you, what business benefits would accrue from a successful EDC project?	This question offers insights from a user perspective around a specific use case for the data catalogue's utilisation.
5. What factors would mitigate the success of an EDC project?	It is expected to draw out any key lessons learnt from the journey to date.

Chapter 4. Methodology

4.1 Choosing a case research approach

The analysis of the specific benefits and issues in the development of an enterprise data catalogue is an exploratory exercise. The anticipated benefits as well as encountered issues are unique to the organisation. Through a single case research approach, the researcher can examine the chosen phenomenon in its natural setting.

4.2 Data collection

A set of questions (see Appendix A) was provided to interviewees at least a week ahead of the interviews, to allow time for preparation and clarification if needed. The questionnaires were composed of 6 questions in total, as described.

4.4.1 Selection of participants

Five people were identified as potential candidates for interviewing. The selection of the participants was based on their involvement and demonstrated interests in the progress of the EDC project, in addition to its findings and outcomes.

Broadly speaking, these five candidates offer a good presentation of different viewpoints. Their roles range from business owner, to project manager, data architect, and business stakeholder. Due to their distinct responsibilities, the researcher believes that their inputs would offer a diversity of opinion to compare.

4.4.2 Data collection process and transcription validation

A Human Ethic Committee (HEC) approval has been granted, and all participants were given an information sheet and participant consent form to sign prior to the interview session.

The interview sessions were scheduled for thirty-five minutes each and were held at the organisation's premises. The researcher conducted the interviews, and audio recordings were taken for each.

Upon completion of the interviews, all of the transcripts were sent back to the interviewees to validate. This also provided them with an opportunity to add or remove any comments from the transcripts. Any feedback provided was consolidated back into the original transcript and only the final versions of the transcripts were used as a base for the data analysis and corresponding findings.

4.4.3 Analysis of Data

The report summarises the key points from each interviewee, based on the results, the report has grouped the comments into four main categories for the data analysis and interpretation section – categorising the Enterprise Data Catalogue, business drivers and measuring success, main Enterprise Data Catalogue Challenges as well as key considerations.

Chapter 5. Summary of Results

Table 1 - Summary of the main points raised by each interviewee.

	Main advantages	Main difficulties	Order	Success	Mitigate Success
Interview A	- Value of the dataset - Orphan materials - Institutional Knowledge - Assets management	- Detective work / understanding - Manageable way of doing it - Publishing mechanism	- Detective work - Recording it in a sensible way with enough data - Using visualisation	Not applicable.	- No funding, no one is interested, no one taking responsibility - Management commitment - Method to do it - Business problem
Interview B	- Increased visibility - Value of the dataset	- Level of maturity - Business engagement	- Business engagement and support - Metadata standard - Method – not big bang	- Relevant - East of use and evolution of it - Relevancy, accuracy of data collected	- Engagement - Data steward responsibilities - Not doing it in silos
Interview C	- Shared vocabulary	- Human factors - Issues on ownership	- By phase - Engagement and advertising, people skills	- Being valued	- Staff turnover - Reorganisation - Legislation change
Interview D	- Consistency - Completeness - Efficiency - People not doing it in silos	- Broader scope - Education / Understanding	- Secure strong sponsorship - Problem statement - Scope - Key user group	- Increased visibility of information	- Sponsorship - Tool
Interview E	- Connections - Strategic VS Operation - Cost-saving - Capability of data stewards	- Data owner / steward - Right people to talk to - Maintenance	- Problem statement - Mandate and funding - Validate the metadata standard	- Coverage - Risk	- Buy-in - Time-frame - Goal

Chapter 6. Analysis and Interpretation

6.1 Categorising the Enterprise Data Catalogue

In order to put the Enterprise Data Catalogue (EDC) into perspective, it is important to first categorise the type of data that is captured by the catalogue. Although there is no direct question asked to interviewees regarding the type of metadata they consider the EDC is, the researcher believes that through the answers of other questions such as benefits and potential usage, enough evidence towards the expected functions it is expected to deliver can be uncovered.

The report has identified and grouped statements from interviewees where they describe the type of metadata or its equivalent functions (see table 2 below). In particular, the data catalogue is described to be an information discovery tool that can be used by both business and IT, and is expected to provide both descriptive and administrative functions of the information resource. In particular, one interviewee described the reason for it is because, "*…Why IT wants metadata and needs metadata is often different to why the business wants and needs metadata…Somehow you have to find a marriage between the two of them…. And make sure you use the language that business understands, and also still useful to IT…*" (Interview A, 2014). It aims to help business people as well as technical users in understanding the data, by providing context to a set of information, such as purpose, description, data stewards, data custodian and update frequency etc., as opposed to documenting data that is for system operation and maintenance i.e. operating system, network components and database servers.

Table 2 Comparison of types of Business Metadata and Interviewee's responses			
Type	Sub-type / Functions	Description	Interviewee Responses
Business Metadata	Administrative	"To manage and administer information resources. Including: - Acquisition information - Rights and reproduction tracking - Location of information - Documentation of legal access requirements"(Gilliland, 2008, p. 9)	**Interview C (2014)** - "*As you build your metadata up across your datasets… you can start recording things around the sensitivity and legal obligations you got with that data.*" **Interview D (2014)** - "*People know where to go…"* **Interview E (2014)** - "*So for the business user of it hopefully, they can more quickly figure out if the data exists. And if it is available to them and who to contact, so connection people and making those connections…*"
	Descriptive	"To describe or identify information resources. Including: - Cataloguing records - Finding aids - Specialized indexes - Hyperlinked relationships between resources"(Gilliland, 2008, p. 9)	**Interview A (2014)** - "*To understand what is in the dataset, its properties, its qualities, and all sorts of other things about it, its relationship to other datasets without actually having to have the dataset at hand…*" **Interview C (2014)** - "*By developing a data catalogue and a shared metadata, it means that people have a way of finding it out that, when people are talking about taxon, and when someone talks about species, they are really talking about the same thing*" **Interview D (2014)** - "*Increased understanding of the information…" "Increased understanding around how data can be used…*"

6.2 Business Drivers and Measuring Success

The creation of an Enterprise Data Catalogue is perceived as a means to address the information-handling problem in the Ministry. In particular, this problem emerged as a result of the recent merger. Like any newly formed organisation, the Ministry has *"…inherited stuff, and maybe the people who maintain that in the previous organisation have left or got other job, and that is no longer their concerns"* (Interview A, 2014). If taking a data as an asset point of view, then the greater danger here is the "*…orphan materials that are actually quite valuable. But just get lost or forgotten about."* (Interview A, 2014).

Subsequently, one of the main business drivers for an Enterprise Data Catalogue is to increase insights and value from the existing data and information. Through the exercises of identifying, authenticating, describing and locating the information resources in a consistent way, it aids in the utilisation of the existing data and information resources by providing visibility of the data within the Ministry. Then the next step can be "*start adding value by looking at how things linked together, and creating new information by understanding those relationships…*" (Interview A, 2014) as well as "*stop people re-inventing the wheel*" (Interview E, 2014).

The Enterprise Data Catalogue is anticipated to provide both short-term and long-term benefits to the Ministry. From a short-term information management perspective, it can help to provide some efficiency gain in some areas, such as "*people being able to quickly see if the data exist before they spend long time deep in the data*" (Interview B, 2014); be able to "*…know who to talk to really understand the data*" (Interview D, 2014); "*better research and better answers to parliamentary questions or OIAs*" (Interview C, 2014); and "*…being more circumspect than how they release that data*" (Interview E, 2014). The development of the Enterprise Data Catalogue is also perceived as "*one of the foundation stones*" (Interview A, 2014) towards the businesses maturity in understanding what data it holds and its potential use of the data. From a long-term information management perspective, the outcome from a well-developed and maintained Enterprise Data Catalogue is the ability to "*operate at a strategic level with the information*" (Interview E, 2014). Using it as a tool in the "*data management space to actually improve how we*

manage the data" (Interview B, 2014) by clearly documenting the business context and relevance of the data, leads to "*better priorities in what ends up in the data warehouse*" (Interview B, 2014). "*Better business engagement…by break down some of the data silos*" (Interview B, 2014) and simplifying the process of finding the right people with the right information. In addition, a consistent approach towards the collection of metadata and its management also helps to discover "*the best place to source that information*" (Interview D, 2014), as a result aiding in the single version of truth through promoting sharing, reuse of knowledge and expertise within the Ministry.

6.2.1 Measuring Success

The approach taken towards the development of the EDC is done in phases, as opposed to a "*big bang*" approach (Interview B, 2014). This is a good and practical way to gain buy-in from the business by demonstrating the value of the Data Catalogue as it develops, as well as publicizing the role of the metadata. All the interviewees commonly acknowledge the continuous improvement aspect in the development of the EDC. One interviewee also explained "*we will not be complete (in terms of identifying stewards and gathering the information) when we publish... in a way because of the publication of it, in itself, it would gather further momentum for it*" (Interview E, 2014). Conversely, it is still important to have a means to evaluate the progress as it develops.

In the responses to the question of how would you measure the success of the Enterprise Data Catalogue, despite the fact that there is a common theme of being successful by getting business value from it, individual measures vary and are valid at different maturity levels of the data catalogue. At the end of the publication phase, the success of the data catalogue can be measured on the *"…increased user awareness of where the data catalogue is located…and increased understanding of how to use data catalogue*" (Interview D, 2014). Only when users become aware of the existence of the data catalogue, the type of information it contains as well as the status of the Data Catalogue through its publication mechanism, can a level of utilisation

be expected. At that point, with some awareness and usage of the data catalogue, then the success of the data catalogue can also be evaluated based on "*people submitting information back to it*" (Interview C, 2014). There is a level of publicity and desire from the business to make their information available to others, through the utilisation of the data catalogue and recognition of its values. Ultimately, there is a level of business ownership to the information that has been submitted to the Data Catalogue. Thereafter, the next stage is the "*ease of use and the evolution of it…continue to evolve as business changes*" (Interview B, 2014). These are only feasible by first demonstrating the value and usability of the data catalogue from previous stages, and some linkages to the business process has developed as it evolves.

In addition, while the "*relevancy, accuracy of the data collected*" (Interview B, 2014) should be the ongoing core measures throughout the entire data catalogue development process. However, without certainty for a level of linkage to the business process, the relevancy and accuracy of any collected data can easily become obsolete as a result of stand-alone processes. Subsequently, only when the relevancy and accuracy of the data is guaranteed to some extent can the data catalogue be used to "*identify high-value public data for release*" (Interview E, 2014). This is achievable by having that linkage to the business process, business ownership and a means to assess the suitability, accuracy and quality of the data for release. As well as a feedback loop mechanism in place to ensure there is a level of synergy between the data that has been released publically and data captured in the data catalogue.

6.3 Main Enterprise Data Catalogue Challenges

The EDC is at an early stage of its development, so the main challenges are around the collecting of required information, and thereafter establishing a robust process for ongoing maintenance of that information. While the Ministry is on the journey to create an Enterprise Data Catalogue, getting business on board to the concept of

metadata initially, has proven to be one of the main challenges, due to different levels of uptake and a lack of clear data guardianship roles.

6.3.1 Lack of uptake

The different level of understanding of the concept of metadata means there is a different level of buy-in towards the purpose and value of the data catalogue. Interviewees indicated that large amounts of effort are required in the promotion and education both initially and on an ongoing basis. In particular, one interviewee described that *"…before they just know they have this data, and they use it for this process and that is that. They don't necessarily see a wider use for it or why we should bother creating a data catalogue."* (Interview A, 2014). Therefore, "*…getting people to understand the concept behind the metadata standards and also the data catalogue. Things like what a dataset is, understanding the difference between a data steward and data custodian…you have to provide people with guidance and education…*" (Interview D, 2014).

In addition, the different level of buy-in also means that for any input required for the data catalogue, a certain amount of investigation and analysis of available information needs to be done beforehand. One interviewee explained this is necessary *"…because they (people in the business) are not that interested, and therefore getting information and keeping it up to date, you can't expect necessarily people to tell you just off their own bat.*" (Interview A, 2014). And also "*depending on where their starting point is, it will depend on how hard it is to work with them to get that buy in*" (Interview E, 2014), as well as getting people to recognise that importance, subsequently to "*be active contributor is a challenge*" (Interview D, 2014).

The different levels of understanding also have implications for the on-going maintenance. One interviewee mentioned that "*for maintenance in particular, you want to make it easy or automating, or even automating reminder that things needs to be refreshed*" (Interview E, 2014). By providing that capability,

then the ideal future state of the maintenance is to "*be owned at a data steward level out in the business* " (Interview E, 2014). However, in order to have that level of ownership, automation or even for the business to respond to any reminder promptly, there is a need to start building on the level of buy-in as well as the rationalisation on the value of the data catalogue in order to acquire that. This also leads to the second challenge, establishing a clear data guardianship responsibility in the Ministry.

6.3.2 Data guardianship

A lack of a clear guardianship from a data perspective means challenges in agreeing on the appropriate caption for the data. Interviewees generally agree that, in order to collect the required information for the data catalogue, there has to be a manual intervention, as *"...there is a certain amount of subjectivity around it, and judgement in deciding what a dataset is…*"(Interview A, 2014). For instance *"…how do business use it, versus how do BI team might view it, versus how the public might view a dataset..."* (Interview B, 2014). Without a clearly defined guardianship or responsibility for a set of information or an information domain, it can be difficult to settle on "*…who makes the call about where is the single source of truth and … a definition of data…"* (Interview B, 2014).

In addition, linking the accuracy and completeness of information in the data catalogue to the relevant data stewards' role, while encouraging collaboration with the cataloguer is also important. By doing so, it helps to secure a level of commitment towards the content in the Data Catalogue. One interviewee mentioned that, "*…often people are so busy doing their real work, they don't have time, even though this would work…"* (Interview B, 2014). Without these measures or that level of formality, the ability to maintain a level of active contribution, therefore preserving the accuracy and relevance of the information in the data catalogue will be compromised. This is especially true when there is an ongoing competition of priorities from other work in the Ministry.

6.4 Key Considerations

In summary, interviewees acknowledge that having a clearly scoped problem statement that explicitly describes the issue that needs to be addressed is the first step. In this case, it is the "*Opacity…you can't see what is out there…that is hindering our ability to be an efficiency organisation*" (Interview A, 2014).

When mixing and matching different standards together to meet that business need, it is important to "*test the standards to make sure people understand it, that the terms make sense, that there is a business alignment*" (Interview B, 2014). Then find a balance between practicalities of the information and the amount of effort required to collect them, have "*enough metadata, but not too much*" (Interview A, 2014). The reason for that is, by not having enough metadata, it would affect the usability of the data catalogue, at the same time, too much metadata means challenges in the gathering of required information and keeping it up to date.

In addition, it is recommended to identify the key user group early on and "*put it in place for the entire project*" (Interview D, 2014). This is raised by interview D as important, because the development of an Enterprise Data Catalogue itself is considered as an "*accumulative endeavour*" (Interview C, 2014), so it helps to raise the awareness of the data catalogue and maintain a level of interest by marketing the usability as it develops. In addition, it also offers regular validation via the user group to ensure "*whatever is delivered provides value to the customer*" (Interview D, 2014).

Chapter 7. Discussion

The creation of an Enterprise Data Catalogue is perceived as a means to address the information-handling problem in the organisation. The issue of a lack of visibility on the data and information the organisation holds leads to challenges in managing it effectively and as an asset. The data catalogue is to carry out the role of business metadata, by providing business context for "data valuation and interpretation" (Shankaranarayanan & Even, 2006, p. 90). Therefore, it helps to address the information-handling problem by documenting then preserving the relevance of the information resources and acting as an information discovery tool to facilitate the "data exchange and sharing" (Owens, 2008, p. 372) in the organization.

Having a clearly scoped problem statement is the foundation for the selection of best-fit metadata elements. Depending on the intended use of metadata and its audience, suitable elements vary. Chagoya (2010, p. 258) suggested it is important to "keep things simple and lean" for the metadata elements, and "only data that is relevant and used should be collected" (Wilson, 2007, p. 17). If data is not relevant, then resources are wasted on its creation and maintenance. On the other hand, if collected data is not used, then it will eventually become obsolete and add no value. To find a balance between the required effort and its usage, an assessment can be done on "how the end user currently acquires the information they need" (Chagoya, 2010, p. 258), and subsequently the usefulness of the metadata elements.

A well-developed and maintained Data Catalogue offers both short-term and long-term benefits to the organisation. The short-term benefits are around efficiency gains in time to answer by locating the right people with the right information faster (Hunter, 2003), re-use and re-purposing of existing information in the organisation by understanding the current information linkage. The long-term benefits for doing so are the ability to operate at a strategic level with its information including better prioritisation, better business engagement, reuse of knowledge and expertise within the organisation and to "manage data as an organisational asset" (Kumar & Palvia, 2001, p. 154).

Conversely, metadata management is a hard sell, as the value of metadata accrues over time. Despite the fact that metadata is an essential part of information management. "Securing organisation wide support is typically the greatest challenge in any successful

metadata implementation" (Shankaranarayanan & Even, 2006, p. 94). That is also reflected in the current level of active contribution to the data catalogue in the Ministry. The different level of understanding towards the concept of metadata also means different levels of uptake towards its value and justification for doing so. Subsequently, it means challenges in securing any long-term commitment from senior-management on the work and resource required. Owens (2008, p. 375) recommended, "The best argument for an enterprise approach is to demonstrate real value". So the publication of a work in progress Data Catalogue in the Ministry is considered as one of the practical steps towards achieving this. Furthermore, "maintaining a constant marketing-style communication program to remind people what metadata is for" (Owens, 2008, p. 375) is also needed in the Ministry to build and maintain the level of understanding towards the concept of metadata and its role in information management.

Incorporating the principle of data stewardship in managing the components of metadata mapping and its management is also important. It helps to safeguard the relevance of metadata management from any organisation's structure, staff or activities changes, therefore reducing the risk and impact on metadata's generation and transformation process (Kurth et al., 2004). By incorporating the principle of data stewardship in the metadata management, it also acknowledges the role of metadata in information management and treats the metadata management as an ongoing accumulative practice, as opposed to a sporadic exercise.

Furthermore, to reach the future state of autonomy in maintenance, particularly for the accuracy and completeness of the Data Catalogue, entries should be owned at a data steward level. There is a decision that needs to be made on the level of integration from the Enterprise Metadata Standard to the business process, system or project (Owens, 2008). The level of integration may vary depending on the nature of the organisation and its maturity towards the metadata concept, however developing a metadata process that promotes sharing and reuse among staff who create and manage information resources can also assist to grow the level of integration and effectiveness of metadata practises (Sun, 2008).

Chapter 8. Study Limitations and Implications

8.1 Study Limitations

There are two main limitations on this case study and its findings. First of all, the researcher is heavily involved in the development of the Enterprise Data Catalogue in the Ministry. Although extra effort is made in order to keep an objective view of the data obtained, it may still implicate on the objectivity involved in interpreting the responses. Secondly, the research was conducted with only five participants in total. The researcher believes the data is derived from a good representation of different viewpoints since the participants' roles are varied, however the sample size may have limited its findings.

8.2 Implications for further research

In addition to the examination of the rationality for taking an enterprise approach towards metadata's development, including its benefits and challenges involved. Subsequent research could evaluate following:

- The feasibility of the level of integration of an Enterprise metadata standard to its business process, systems and project.
- The effect of the data stewardship concept on ongoing metadata management.
- The effectiveness of the Enterprise Data Catalogue to the Ministry's decision-making process.

Chapter 9. Conclusion

Metadata is an essential part of information management. From a theory perspective, the benefits of an enterprise wide approach towards the design of a metadata standard can lead to enhanced search capabilities, expanded use of data, and support for master data management. On the other hand, such a wide approach also comes with its challenges including, identifying the appropriate metadata standard, metadata management process, as well as finding a best-fit technology to utilise.

In practise, the research identifies that the enhanced search capabilities and expanded use of data are the main business drivers for development of an Enterprise Data Catalogue within the Ministry. Particularly, the Enterprise Data Catalogue is anticipated to offer both short-term and long-term benefits to the Ministry. The short-term benefits are around efficiency in time to answer, re-use and re-purposing of existing information. The long-term benefits are the ability to operate at a strategic level with its information including better prioritisation, better business engagement, reuse of knowledge and expertise within the organisation. On the other hand, different levels of understanding towards the concept of metadata within the Ministry, also leads to challenges in the initial uptake, then securing and maintaining an active level of contribution from business. In addition, a lack of a clear guardianship from a data perspective means challenges in the ongoing maintenance and agreeing on any changes in the data captions.

The report recommended maintaining a constant marketing style of communication in promoting the value and understanding of metadata is the key to acquiring buy in and long-term commitment. In addition, it is recommended to incorporate the principle of data stewardship in managing the component of metadata mapping and its management. This ensures the relevance of metadata management is safeguarded and reduces the risk of rework as well as impact on metadata management from any changes in the organisations structure, staff or activity.

References

Baca, M. (2008). *Introduction to metadata.* Los Angeles, CA: Getty Research Institute.

Chagoya, F. (2010). Metadata: Principles, practical application, best practices, optimization and workflow. *Journal of Digital Asset Management, 6*(5), 257-261. doi: 10.1057/dam.2010.27

Franks, P., & Kunde, N. (2006). Why metadata matters. *Information Management Journal, 40*(5), 55-58, 60-61.

Gilliland, A. J. (2008). Setting the stage. In T. Gill, A.J. Gilliland, M. Whalen & M.S. Woodley, *Introduction to metadata* (2nd ed.). [Online edition] Retrieved from http://www.getty.edu/research/publications/electronic_publications/intrometadata/setting.pdf

Hamzah, M., & Sobey, A. (2012). The use of business metadata to support decision-making processes. *International Journal of Innovation, Management and Technology, 3*(4), 449. doi: 10.7763/IJIMT.2012.V3.273

Hunter, J. L. (2003). A survey of metadata research for organizing the web. *Library Trends, 52*(2), 318-344.

Inmon, W., O'Neil, B., & Fryman, L. (2008). *Business metadata: Capturing enterprise knowledge.* Burlington, MA: Morgan Kaufmann.

Kumar, A., & Palvia, P. (2001). Key data management issues in a global executive information system. *Industrial Management + Data Systems, 101*(3), 153-164.

Kurth, M., Ruddy, D., & Rupp, N. (2004). Repurposing MARC metadata: Using digital project experience to develop a metadata management design. *Library Hi Tech, 22*(2), 153-165.

Lubas, R. L., Wolfe, R. H. W., & Fleischman, M. (2004). Creating metadata practices for MIT's OPenCourseWare project. *Library Hi Tech, 22*(2), 138-143.

Owens, C. (2008). The practicalities of effective enterprise metadata and taxonomies: Their governance and ethics. *Journal of Digital Asset Management, 4*(6), 372-376. doi: 10.1057/dam.2008.40

Roszkiewicz, R. (2010). Enterprise metadata management: How consolidation simplifies control. *Journal of Digital Asset Management, 6*(5), 291-297. doi: 10.1057/dam.2010.32

Shankaranarayanan, G., & Even, A. (2004). Managing metadata in data warehouse: Pitfalls and possibilities. *Communications of the Association for Information Systems, 14(1)*, 247-274.

Shankaranarayanan, G., & Even, A. (2006). The metadata enigma. *Association for Computing Machinery.Communications of the ACM, 49*(2), 88. doi: 10.1145/1113034.1113035

Silvola, R., Jaaskelainen, O., Hanna Kropsu-Vehkapera., & Haapasalo, H. (2011). Managing one master data - challenges and preconditions. *Industrial Management + Data Systems, 111*(1), 146-162. doi: 10.1108/02635571111099776

Stefanov, V., & List, B. (2007). Explaining data warehouse data to business users: A model-based approach to business metadata. *Proceedings of the Fifteenth European Conference on Information Systems, ECIS 2007, St. Gallen, Switzerland, 2007. University of St. Gallen 2007.* (Paper 126). Retrieved from http://aisel.aisnet.org/ecis2007/126

Sun, L. (2008). A metadata manager's role in collaborative projects. *The Electronic Library, 26*(6), 777-789. doi: 10.1108/02640470810921574

Technical Committee ISO/TC 46. (2001). *Information and documentation: Records management. Part1, General.* Geneva, Switzerland: ISO.

Technical Committee ISO/TC 46. (2009). *Information and documentation: Managing metadata for records. Part 2, Conceptual and implementation issues.* Geneva, Switzerland: ISO.

Vander Noot, T.,J. (1998). Libraries, records management data processing: An information handling field. *ARMA Records Management Quarterly, 32*(4), 22-26.

Westbrooks, E. L. (2004). Distributing and synchronizing heterogeneous metadata in geospatial information repositories for access. Retrieved from http://digitalcommons.unl.edu/libraryscience/165/

Westbrooks, E. L. (2005). Remarks on metadata management. *OCLC Systems and Services, 21*(1), 5-7.

Wilson, A. J. (2007). Toward releasing the metadata bottleneck: A baseline evaluation of contributor-supplied metadata. *Library Resources & Technical Services, 51*(1), 16-28.

Wu, Y. B., & Li, Q. (2008). Document keyphrases as subject metadata: Incorporating document key concepts in search results. *Information Retrieval, 11*(3), 229-249. doi: 10.1007/s10791-008-9044-1

Zeng, M. L. (1999). Metadata elements for object description and representation: A case report from a digitized historical fashion collection project. *Journal of the American Society for Information Science, 50*(13), 1193-1208.

Appendix A - Interview Questions

1. Explain what are the main advantages of a corporate approach to metadata's creation, development and maintenance?
2. What are the main difficulties one may expect when taking a corporate or enterprise approach in creating, developing and maintenance of business metadata?
3. Describe the order you would follow to implement a successful EDC?
4. How would you measure the success of an EDC project?
5. What factors would militate against success of an EDC project?
6. For you, what business benefits would accrue from a successful EDC?