eChallenges e-2010 Conference Proceedings
Paul Cunningham and Miriam Cunningham (Eds)
IIMC International Information Management Corporation, 2010
ISBN: 978-1-905824-20-5

Information Governance and Metadata Strategies as a Basis for Cross-sector e-Services

Terje GRIMSTAD[1], Per MYRSETH[2]
[1]*Karde AS, P.O. Box 69 Tåsen, Oslo, N-0801, Norway*
Tel: +47 90844023, Email: terje.grimstad@karde.no
[2]*Det Norske Veritas, Veritasveien 1, Høvik, N-1322, Norway*
Tel: +47 41855885, Email: per.myrseth@dnv.com

Abstract: Information governance and systematic work with metadata and semantics are important elements of the implementation of an open, transparent, accessible, accountable, user-friendly and service-oriented public sector. Top management commitment is crucial in order to achieve necessary attention and sufficient budgets. Management needs to be aware of metadata and semantics as important enablers for the goals set forth in strategies and requirements from ministries. Documentation of economic potential and cost savings will help to get attention among decision makers, but today, few trustworthy sources are available. The development of cross-sector services and the demand for reuse of public service information, both in the public sector itself, but also for commercial services, underpins the importance of well-defined information. Participation in cross-sector e-Services demands the establishment of metadata repositories and ontologies as obligatory parts of the public sector information governance regimes.

1. Introduction

Metadata and semantics are important and necessary elements of the implementation of a collaborative, effective and efficient e-Government. One problem, however, is that these topics, and the necessity to handle them in a structured and systematic way, are scarcely understood by top management with decisive power. Top management commitment becomes a crucial success factor. There is a need to convey a convincing, trustworthy and understandable vision and mission statements in order to get the necessary management attention and financial support for metadata and semantic exercises. The effects of systematic use of metadata across the public sector must be expressed in a language that is understood by top management, with qualitative and quantitative measures. The effects, which are both internal and external to the public organizations, have to be clearly documented and communicated.

The semantics are often hidden in source code, in systems, in laws and regulations, in the organizational structures and in the tacit knowledge of collaborating colleagues. Therefore, semantic problems have traditionally been satisfactorily dealt with by bi-lateral agreements between the ICT-staffs of the collaborating parties.

In most countries it is an important goal to provide an open, transparent, accessible, accountable, user-friendly and service-oriented public sector. A crucial requirement to obtain this is better collaboration between public entities, and more seamless solutions between businesses, citizens and the public sector. As a consequence, more information exchange becomes necessary. Bi-lateral agreements become unmanageable when the number of collaborative cross-sector efforts increases. A new paradigm is called for.

Metadata and semantics are crucial building bricks in an information governance regime. Also, they constitute instruments to obtain better collaboration both within public organisations, between public organisations and between the public sector and its users.

The main objective of this paper is to show that information governance and adequate metadata strategies are important issues for the implementation of an open, transparent, accessible, accountable, user-friendly and service-oriented public sector. Other objectives are to provide guidelines for the content of the metadata strategy and elements of a "selling story", based on best practices and experiences from interoperability research in the Norwegian public sector. The selling story is aimed at top management and meant to increase their understanding of the role of metadata and semantics in the development of eGovernment.

2. State of the Art and Best Practices

2.1 Norwegian Initiatives

Norway has a well-organized public sector with several infrastructure services already in place, or under development. The following initiatives and activities have been used as background and inspiration for this paper:

1. Altinn (www.altinn.no/en) is a service through which citizens and businesses can report information to public authorities.
2. MyPage (www.norge.no/minside) is a portal through which services from different public bodies are made available to the citizens.
3. SERES is a service in development run by the Brønnøysund Register Centre, the goal of which is to provide a national metadata register and repository. SERES is treated in some detail in the next subchapter.
4. eDialogues is a concept for the implementation of cross-sector services to citizens and businesses.
5. Statistics Norway has for several years worked systematically with metadata and obtained several gains. Top management adopted a metadata strategy in early 2005 [1].
6. Two preliminary efforts to provide requirements for a national metadata strategy.
7. Ongoing work to define a national metadata strategy.

2.2 Metadata and Metadata Strategies

Initiatives for describing metadata are not new. Attempts to describe semantics in data exchange have traditionally had many flavours. In the following some examples are given.

UN/CEFACT

UN/CEFACT initiatives, starting back in late 1980s [2], define international electronic exchange formats and semantics for orders, invoices, custom declarations etc. Organisations like OASIS Universal Business Language (UBL) [3] and CEN [4] are also involved in the standardisation of these kinds of business documents. A national subset of invoices and credit notes has been made for the Norwegian public sector. It is a goal that the Norwegian public sector shall be able to receive electronic invoices and credit notes by July 2011. By June 2012 the municipalities shall be enabled, and it is a requirement that the private sector shall send these documents in electronic formats [5] by the dates specified. In February 2005, Denmark made it obligatory that all invoices sent to the public sector should be in electronic format [6].

Health Level 7

In the health sector, the non profit organisation HL7 was established in 1987. It provides a framework and related standards for the exchange, integration, sharing, and retrieval of electronic health information [7]. In Norway the Norwegian Centre for Informatics in

Health and Social Care [8] is the national body for aligning the HL7 standards with national legislation and to assure pragmatic adoption to achieve network effects.

European Business Register

The European Business Register [9] gives online and easy access to reliable information about 20 million companies from 19 different European countries. The information is up to date from each country's official register. The national bodies like The Brønnøysund Register Centre in Norway, The Swedish Companies Registration Office (Bolagsverket) and The Danish Commerce and Companies Agency (Erhvervs- og Selskabsstyrelsen) are three of the registration authorities supplying these data. In a press release from November 2009, the EU Commission pin-points the importance of interoperability between business Registers [10]:

"*The current financial crisis highlighted once again the importance of transparency across the financial markets. .... Business registers play an important role in ensuring transparency and legal certainty in Europe. ... Co-operation between business registers remains voluntary and does not seem to be sufficient to achieve certainty in all cross-border legal procedures and to increase transparency in the Single market.*"

A cross sector e-Service on a national level in Norway is "Collaborative enterprise information update" ("Samordnet registermelding" [11]). This electronic reporting obligation distributes updated enterprise information to six different Norwegian public registers holding Norwegian enterprise data, and validates the input data according to the rules of the destination registers.

The above mentioned initiatives all use metadata strategies to secure success. They all have one or more international standardisation/best practice activities to ensure the establishment of (i) terms; (ii) exchange models of data; (iii) process models describing sequences of data flow, error handling; and (iv) technical protocols for sending, receiving, security mechanisms, message envelope design etc.

All the standards and recommendations are tailored nationally to comply with national legislation, the installed base of e-Government services and the departments' internal solutions and needs. A challenge, though, is that the national registers often were established before the international standard effort was finished and they were based on the current national legislation. Because of this, EU is in a situation where the national registers are only partially interoperable when it comes to (i) legislation, (ii) content in the registers and (iii) information governance regime for the registers. This is a common challenge in eGovernment service interoperability, where services are established and funded for one purpose, but later in their lifespan broaden their purpose and have to adapt to new legislation or other services' needs.

The public sector sets the rules of the game, which can be described as a "carrot and stick": If you don't play by the rules you are not able to perform electronic business with the public sector. E.g., when the public sector does not accept paper invoices, the incentives to play by the rules on the electronic arena are very strong. Regarding the European Business Register, only authorized national bodies supply data, and businesses have a strong economic self-interest in validating their business partners, suppliers and customers.

SERES

The Semantic Register for Electronic Services [12] is a cross sector e-Government service for metadata. The Brønnøysund Register Centre, which is in charge of SERES, has established an organisation, a methodology and supporting software tools enabling the Norwegian public sector to collaborate on making, maintaining and publishing metadata. The technical solution helps the users in the design and generation of message payload declaration as XML Schemas [13]. These XML Schemas are used to perform both silo e-Government services and cross sector e-Government services in e.g. Altinn. So far the metadata kept in SERES is fit to handle silo e-Government services in Altinn. SERES has

the functionality to facilitate modeling of different domains within the public sector. The fundament for models in SERES is based on UML [14] and the meta-models for UML-like MOF [15]. Further, models at different abstraction levels and versioning of the metadata in the SERES metadata repository is based on a combination of UML and ISO 11179 [16].

The success of SERES as a component in a national metadata strategy is linked to its adoption rate and its ability to meet the information governance needs of its public sector users.

2.3 Process and Information Models

One of the findings in a study by Hellman [17] on barriers to interoperability is that interoperability must be based on good understanding of business processes and information, and models of these. The study also reveals that the descriptions of business processes and information in the Norwegian public sector are not satisfying. Also, there is a lack of competence and understanding at all levels in the public organisations.

Information governance is a key success factor for effective and efficient government. Based on challenges in the public sector, we assert that the principles presented by Khatri and Brown in "Designing Data Governance" [18] are relevant to the public sector. Khatri and Brown [18] divide information governance in five sub-domains: 1) Information principles – clarifying the role of information as an asset; 2) Information quality – establishing the requirements of intended use of data; 3) Metadata – establishing the semantics or "content" of information so that it is interpretable by the users; 4) Information access – specifying access requirements of information; and 5) Information lifecycle – determining the definition, production, retention and retirement of information.

The public sector can be regarded as an organism which is managed by laws and regulations, and which operates on information. There is a dependency triangle between the (i) legal context in which an enterprise operates, (ii) the actual operations and type of business performed and (iii) the information assets that an enterprise utilizes to perform these operations. The dependency is such that if e.g., the legal context evolves, the enterprise information governance regime has to adapt to new situation and maybe the operations also will have to change. Further, if an enterprise starts operating a new type of business or starts operating in a new country, the legal context changes, which in turn may impact the information governance regime etc. This triangle is based on our findings in our cases studies and in the reports "Compliance Work Package" [19] covering legal compliance in information governance in EU and "Semantic Technologies in Information Governance" [20].

3. Developments

3.1 Information Governance and Metadata Strategies

As stated above, the public sector can be regarded as an organism. Therefore, there is a need to enhance the sub-domains from Khatri and Brown [18] with a sixth domain, namely Information Compliance in the context of laws and regulations.

A metadata strategy should address all the six domains suggested above, and act as the guideline for information governance. As a result of the implementation of the strategy, several effects, both internal and external to the public organisation, will become visible. These effects are described in the next chapter.

Information governance focuses on who holds the decision rights on which topics, while management is about implementing the decisions [18]. There is no complete list of which topics to decide upon, but the dependency triangle may help. The dependency

triangle mentioned earlier illustrates the dependency between (i) legal context, (ii) the actual operations and (iii) the information assets.

3.2 Different Uses of Metadata

In this article we use the term metadata meaning "data about other data" or "data about data", as it is defined in the Reference Model for an Open Archival Information System (OAIS) [21] and in ISO 11179 [16]. Metadata and ontologies may serve many purposes, but we limit our focus to metadata with the purpose of identifying and defining the meaning of data in cross sector e-Government services.

There is a need for methodologies and tools to formalize the metadata into an asset that can be maintained and governed. When the metadata is formalized, maintainable and shareable new ways of utilizing metadata arise. Some of these are:

- Internal to a department metadata will help to get an overview of the current situation when it comes to available data and the meaning of the data. Well defined metadata makes it easier to maintain a portfolio of interconnected internal systems.
- A metadata repository containing several departments' metadata will be of great help when metadata harmonisation processes are carried out.
- Information exchange that crosses system or juridical boundaries uses metadata and exchange models as a contract for the exchange protocol.
- In participation in cross sector e-Government services, e.g. ability to map/align data to standardized models, to other actors' models, to alternative representational formats and ability to adapt to new regulation or collaboration.
- In order to measure data quality there is a need to measure actual data according to some rules. The metadata can be the main part of the rules.
- For Business intelligence and statistics metadata is used as a description of the meaning and the identifiers used.
- Open public data/ linked open data. Uses metadata as a description of the meaning and the identifiers used.
- Development of legislation. The metadata repository gives an overview of existing terms and their usage. This knowledge is very important when designing new or checking the consistency of existing legislation.
- Simulating the impact of changes of regulation. Well designed metadata can help simulate the effect of changes in e.g. tax regulation. Metadata described in an ontology language may help answer questions like "what will the national tax income be if the definition of income is changed from x to y".

3.3 The Importance of Public Sector Strategies and Letters of Instruction

In order to obtain necessary management attention and commitment, experiences from our research indicate strongly that information governance and metadata strategies have to be bound to public sector strategies and goals set by the ministries when the yearly budget for the public organisation is allocated and formulated in a so-called "letter of instruction".

Strategies often contain statements regarding different characteristics of the public sector. Some examples are:

- About **Openness**: The organisation should be the preferred source for information; The organisation must protect privacy.
- About **Innovation**: The organisation should provide Better services for citizens and businesses.

- About a **User friendly public sector**: The organisation must take the initiative and participate in cross-sector service development; The organisation should contribute to service developments in other sectors.
- About **Quality culture**: The organisation should improve existing services; The organisation should work more effectively and efficiently.
- About **Competence culture**: The organisation should develop competence in a systematic way; The organisation should transform individual knowledge to common knowledge; The organisation should build, use and share knowledge

In the letter of instruction from the ministries to the different organisations, the organisations are asked to position themselves according to development trends in the society, e.g.

- It must provide accountable and efficient governance of a large amount of information.
- It must provide better communication with users as a result of new ICT.
- It must position itself to support increased demand and need for collaboration with other public entities.
- It must position itself to increased quality demands on services and products.
- It must position itself to increased demand on competence and management.
- It must position itself to increased demand and requirements for rules developments.

These requirements all ask for a high level of semantic consciousness, information governance and a metadata strategy. The communication towards top management could be based on pictures like the one in figure 1. The idea of this visualisation is to show the relative importance of semantic consciousness for each of the requirements. The outer circle indicates the level which can be obtained if the organisation works systematically with semantic issues. The assessment of the level values are based on both qualitative and quantitative considerations. E.g. for "accountable and efficient governance of large amounts of data", the importance of systematic work with semantics is assessed to a level 5 out of 6.

The inner circle indicates the level of semantic support provided by the present methodologies and system portfolio. The figure shows that there is a gap to close in order to fulfil both strategies and demands from the ministry. To close the gap, which is a necessity because the organisation is measured on the degree of fulfilment, requires management commitment, budget and competence development. Other measures are also necessary in order to reach level 6, e.g. enterprise models, efficient and effective business processes, systems support, project management. Semantics is not everything.

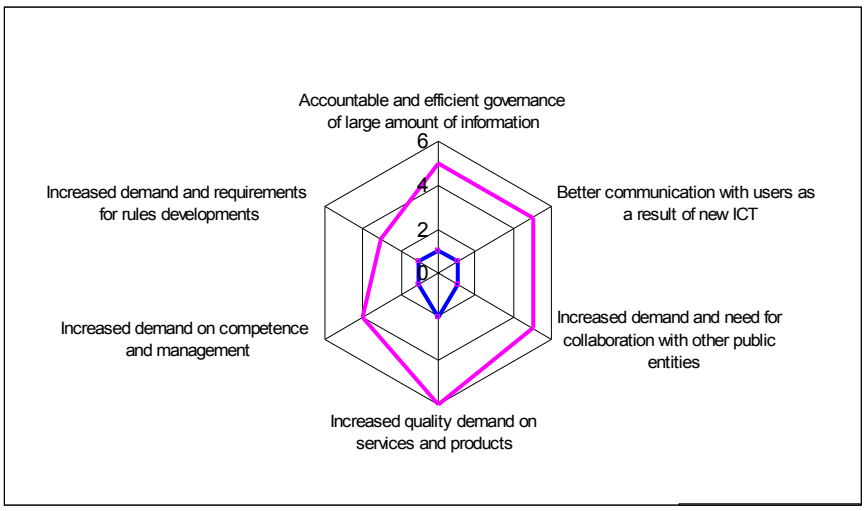

Figure 1: Gap Between Challenges and Present Situation

4. Effects of Information Governance and Metadata Strategies

Statistics Norway has published several internal effects of their metadata strategy [1]. They claim that the metadata strategy is a cornerstone in their ICT-strategy. Some of the effects are better quality statistics, avoidance of double work in the production, increased reuse, increased ability to cooperate between departments, identification of incompatible definitions of the same term, more robustness in relation to change of personnel, reduced demand for user support and an increased ability in the organisation to change.

Our research has identified additional effects. We classify these effects as internal or external to the organisation. Internal effects can be summarised as:

1. By working with Information governance in a structured manner, business becomes deeply involved in the definition of concepts. This in turn leads to better alignment between the business processes and the ICT-solutions and the ability for businesses to develop services with lower degree of ICT.
2. Individual knowledge is transformed to common knowledge. This is due to better documentation, i.e., overview of information, systems and processes.
3. Due to better documentation, the organization becomes more independent of specific resources and more robust to the exchange of personell.
4. Less production errors, and as a side effect, less negative attention in the media.
5. More efficient service development, more efficient systems development and maintenance, easier adaptation of systems to new rules and legal constraints.
6. As a consequence of all effects, the competence and capacity of the staff increases without employing more people. The ability for innovation increases.

External effects can be summarised as:

1. The publication of own information in such ways that it can be reused both for cross-sector services and for commercial services.
2. Avoidance of double reporting obligations for citizens and businesses.
3. More effective and efficient cross-sector service development.
4. Improved implementation of rule of law principles.
5. Improved interoperability.

5. Conclusions

In this paper we have presented arguments for the importance of information governance in the public sector and the role of metadata and semantics. Our observation is that management attention and comprehension is crucial for the implementation of a sufficient information governance regime. Management needs to be aware of metadata and semantics as a crucial enabler for the goals set forth in strategies and requirements from ministries. Furthermore, other effects are also of value and must be communicated. Further work is needed on quantitative effects of good use of metadata.

- What is the cost of double reporting? To what degree will double reporting be reduced?
- What is the cost of production errors? How many percent will production errors be reduced – 10-20-50%?
- What is the cost of training a new employee? Ho many percent will this cost be reduced if the systems are well documented?
- What is the cost of negative press?
- To what degree can system development and maintenance costs be reduced?
- To what degree will service development costs be reduced? For internal services, for cross sector services?

Recommendations for the public sector are:

1. Increase the understanding of national and international metadata strategies.

2. Visualize important elements of information governance and metadata strategies so that they are understood by top management.
3. Visualize the importance of metadata strategies for the re-use of Public Sector Information, e.g. Review of Directive 2003/98/EC – [22], which claims that PSI has the potential for an immense commercial value.
4. Predict effects of systematic work with metadata and semantics.
5. Visualize the necessity of metadata strategies for the development of cross-sector services.
6. Visualize the need for a new or existing public agency with the role of operating a national metadata service with a clear mandate from the ministries.

Acknowledgements

This research was partially funded by the Semicolon-project [23], supported by the Norwegian Research Council, contract no. 183260.

References

[1] Meatadata strategy in Statistics Norway. June 2005. SSB report 8/2005.
[2] URL: http://en.wikipedia.org/wiki/UN_CEFACT_TBG5#History_of_EDIFACT_and_UN_CEFACT_TBG5
[3] URL: www.oasis-open.org/committees/ubl
[4] URL: www.cen.eu/cwa/bii/specs
[5] Implementation guide, invoice and credit note. (In norwegian only) The Agency for Public Management and eGovernment (DIFI - Direktoratet for forvaltning og IKT), 4 May 2010
[6] URL: www.epractice.eu/en/document/288205
[7] URL: http://en.wikipedia.org/wiki/Health_Level_7
[8] URL: www.kith.no
[9] URL: www.ebr.org
[10] Getting better access to company information: Commission consults on the interconnection of business registers. EU Commission, Press release 5 November 2009. URL:http://europa.eu/rapid/pressReleasesAction.do?reference=IP/09/1677&format=HTML&aged=0&language=EN&guiLanguage=en
[11] URL: www.brreg.no/blanketter/samordnet.html
[12] SERES, Semantic Register for electronic Services. URL: www.brreg.no/samordning/semantikk/index.html
[13] XML Schema. URL: www.w3.org/TR/2004/REC-xmlschema-0-20041028/
[14] UML, Unified Modeling Language. URL: www.uml.org
[15] MOF Object Management Group's MetaObject Facility. URL: www.omg.org/mof
[16] ISO 11179, Information technology Metadata Registries
[17] Organizational Barriers to Interoperability. By R. Hellman. Norwegian Case Stydy. H. J. Scholl, M. Janssen, R. Traunmüller, M. A. Wimmer (Eds.): Electronic Government: Proceedings of ongoing research and projects of EGOV 09. 8th International Conference, EGOV 2009. Trauner Druck: Linz, Schriftenreihe Informatik # 30, 2009. Pp. 182-189
[18] Vijay Khatri and Carol V. Brown, Designing Data Governance, Communications of the ACM, January 2010, Vol. 53, No. 1, pp 148-152
[19] Compliance work package, State-of-the-Art. Longrec report, Hannelore Dekeyser, September 2008. URL:http://www.longrec.com/Intranet/ResearchResults/StateOfTheArt/LongRecCompliance_StateOfTheArt_2008.pdf
[20] Semantic Technolgies in Information Governance. Longrec Understand Research Report, April 2010. Veronika Haderlein, Per Myrseth and Olga Cerrato.
[21] Reference Model for an Open Archival Information System (OAIS). Januar 2002, Consultative Committee for Space Data Systems.
[22] COMMISSION STAFF WORKING DOCUMENT, Brussels, 7.5.2009, on the re-use of Public Sector Information. Review of Directive 2003/98/EC. URL: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0212:FIN:EN:HTML
[23] Semicolon project. URL: www.semicolon.no