State-of-the-Art Assessment on the Implementations of International Core Data Models for Public Administrations

Mathias Van Compernolle[1]
mathias.vancompernolle@ugent.be

Laurens De Vocht[2]
laurens.devocht@ugent.be

Stijn Goedertier[3]
stijn.goedertier@be.pwc.com

Nikolaos Loutas[3]
nikolaos.loutas@be.pwc.com

Ruben Verborgh[2]
ruben.verborgh@ugent.be

Erik Mannens[2]
erik.mannens@ugent.be

Peter Mechant[1]
peter.mechant@ugent.be

Vassilios Peristeras[3]
vassilios.peristeras@ec.europa.eu

Rik Van de Walle[2]
rik.vandewalle@ugent.be

[1] iMinds – MICT
Ghent University
Ghent, Belgium

[2] iMinds – Multimedia Lab
Ghent University
Ghent, Belgium

[3] ISA - DIGIT
European Commission
Brussels, Belgium

ABSTRACT

Public administrations are often still organised in vertical, closed silos. The lack of common data standards (common data models and reference data) for exchanging information between administrations in a cross-domain and/or cross-border setting stands in the way of digital public services and automated flow of information between public administrations. Core data models address this issue, but are often created within the closed environment of a country or region and within one policy domain. A lack of insight exists in understanding and managing the life-cycle of these initiatives on public administration information systems for data modelling and data exchange. In this paper, we outline state-of-the-art implementations and vocabularies linked to the core data models. In particular we inventoried and selected existing core data models and identified tendencies in current practices based on the criteria c*reation, use, maintenance* and *coordination*. Based on the analysis, this survey suggest research directions for policy and information management studies pointing to best practices regarding core data model implementations and their role in linking isolated data silos within a cross-country context. Finally we highlight the differences in their coordination and maintenance, depending on the state of creation and use.

CCS Concepts

• **Information systems~Information integration** • Information systems~Resource Description Framework (RDF) • **Social and professional topics~Governmental regulations** • **Applied computing~Computing in government**

ICEGOV '15-16, March 01-03, 2016, Montevideo, Uruguay
© 2016 ACM. ISBN 978-1-4503-3640-6/16/03…$15.00
DOI: http://dx.doi.org/10.1145/2910019.2910027

Keywords

Core Data Models; Interoperability; Lifecycle; Inventory; eGovernment; Governance

1. INTRODUCTION

A cross-domain flow of information is necessary to implement a whole-of-government approach in which public administrations are organised as formal and informal networks instead of isolated silos. From an eGovernment point-of-view, such coordinated networks enhance efficiency by reducing duplications of processes and procedures in programme management and service delivery, e.g. by sharing a common understanding on how to exchange and understand information moving from one functional area to another. *Core data models* address this issue.

Furthermore, there is lack of overview of how these data models came to their current form, and the nuances among their common similar goal.

One of these goals could be achieving better information exchange across administrations and country borders about people, organisations, locations and public services. This could create business value for enterprises by reducing the red tape and creating increased semantic interoperability [1]. The core data models and their vocabularies provide the building blocks to make these objectives happen.

A core data model is one of the key building blocks for scalable assimilation of information from diverse data sources, next to methods for publishing and disclosing information. This is because they emerge as an accepted data model that expresses basic concepts that are common across a variety of domains and provide the basis for specialization into domain-specific concepts and vocabularies, and thereby facilitate well-defined compatibility between local or region-specific models. Moreover, they come forth as simplified, reusable and extensible data models that capture the fundamental characteristics of a data-entity in a context-neutral way[1].

The idea of core data models for public administrations is closely related to the idea of *once-only approaches* –whereby public administrations don't request information from citizens and businesses that already have been provided in another context, increasing government effectiveness and efficiency, and administrative burden reduction [2] [3]. The overall goal is to provide consistency, standardization and improved efficiency in information exchange and to enable public administrations to execute policies which cut across several functional areas and different domains from the information management point-of-view (e.g. development, maritime policies, and environmental protection). The role of central governments is to support and monitor these information management processes that should lead to better interoperability as they entail an important public good. For example, providing direct feedback such as an interoperability score when a new dataset has been published can help the adoption of the available vocabularies [4].

One of the main actions of the Interoperability Solutions for Public Administrations (ISA) Programme was to promote semantic interoperability among the European Union Member States. Under this action four core data models have been developed so far in an open and inclusive process[2]: Core Person, Registered Organization, Core Location and Core Public service. The use of core data models is part of a solution to achieve seamless cross-domain information flows in the public sector. The international Community of Practice on Core Data Models, a network of representatives in EU Member States, define 'core data models' as "reusable data models that are defined, managed, promoted and maintained centrally to facilitate interoperability across different systems, applications and domains"[3].

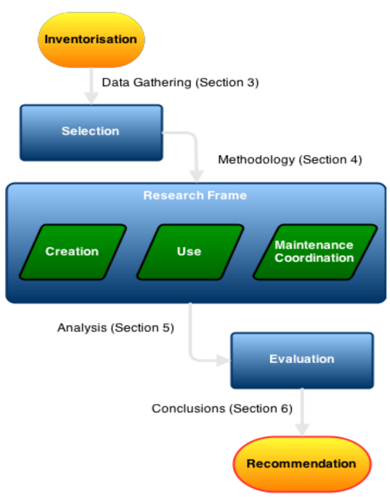

Figure 1. Schematic outline of the survey paper

This paper describes recurring tendencies in existing implementations of core data models. In particular we will focus on (i) their origin, (ii) how they are created and (ii) how they are administered. We formulate recommendations regarding future implementations of core data models and the analysis of their impact on interoperability. We describe and evaluate the current state of the art on core data models from a global perspective, based on *a priori* defined criteria. **Figure 1** outlines the process we used throughout the paper.

First, we create an inventory of projects in which new core data models are developed or existing models are adjusted for reuse. We select core model implementations and create an overview of encountered vocabularies. Next, we describe the methodology and identify the criteria for our analysis. These criteria were discussed during the core data model workshop organised by ISA in Brussels, November 2014[4]. These criteria resemble the life cycle of core data models. Additional data was gathered by desk research and document analysis. These results are discussed in section 5 and lead to recommendations regarding existing and future implementations of core data models.

2. RELATED WORK

E-government still is highly fragmented in continental Europe. Bovalis et al [5] explain this by the fact that governmental institutions, unlike private sector bodies, are faced with unique obstacles, such as complex organisations, lack of profit incentives, differences in data protection and legal constraints within the different member states. To the best of our knowledge, there is no previous work that describes the motivations for creating, using and the maintenance and coordination aspects on core data models from a global point-of-view. Our paper bridges this gap, by analysing existing efforts on core data models, especially those designed for the direct use in public administrations. Below, we will introduce relevant related terminology and we walk through the different interoperability strategies and models around core data models. Hereby we look at the generic characteristics - less topic- or content-driven - of how data is structured and what the modelling process behind it is.

2.1 Core Data Models and Linked Data

Linked Data is a method of publishing structured data so that it can be interlinked by adding semantics to data and its domain model. It builds on standard Web technologies such as "*Uniform Resource Identifier"* (URI) and the "*Resource Description Framework"* (RDF). An URI is a string of characters used to identify a name of a resource. Using core data models and as such describing data in RDF, which was originally designed as a meta-data model - has become a general method for conceptual describing or modelling information that is implemented in web resources.

Linked Data allows data from different sources to be connected and queried. Linked Data deals with the lack of standardization by allowing to cooperate, which enables to bridge local and regional heterogeneities via the flexibility linked data offers. Thus, data integration based on linked data can be considered a way for

[1] EC DG Informatics. eGovernment Core Vocabularies. The SEMIC.EU approach. 2011. *https://joinup.ec.europa.eu/sites/default/files/c1/23/4a/egovernment-core-vocabularies.pdf.* Accessed: 2015-12.

[2]ISA. Improving semantic interoperability in European eGovernment systems. *http://ec.europa.eu/isa/actions/01-trusted-information-exchange/1-1action_en.htm.* Accessed: 2015-12.

[3]Kotoglou, S. Community of Practice on Core Data Models. *https://joinup.ec.europa.eu/community/semic/document/community-practice-core-data-models.* Accessed: 2015-12.

[4] Workshop Core data models for public administrations – 12 november 2014, Brussels. *http://joinup.ec.europa.eu/site/core_vocabularies/2014-11-12_Core_data_models_workshop/2014-11-12_Core_data_models_workshop_Report_v1.00.pdf* Accessed 2015-12

standardization to have an effect in early stages of data model developments and before the standard is completed or perfected [6].

Converting data to a set of RDF triples and linking them to another set of triples does not necessarily make the data more (re)usable or interoperable. We use the term *vocabulary* here to refer to the aspect of semantics and to describe schemas, ontologies, taxonomies, terminologies etc. While there is a risk for over-engineering, a good vocabulary should restrict potential interpretations of the used classes and roles towards their intended meaning [7].

Such identification enables interaction with representations of the resource over the web. Schemes specifying a concrete syntax and associated protocols define each URI. The most common form of URI is the uniform resource locator (URL), frequently referred to informally as a web address. If we want a broad adoption of Linked Data, which describes a method of publishing structured data so that it can be interlinked, the barrier to conform to the Linked Data principles need to be as low as possible. One of the Linked Data principles is that URIs should be dereferenceable [8]. Often namespaces are introduced as holders for recurring URIs, in particular their domains and a fixed prefix. Namespaces make it possible to distinguish between identifiers with the exact same name.

Advocating the importance of this process enforces and facilitates linking of data in general. It is a relevant practice in terms of interoperability to tackle the semantic layer separately from the object, syntactic and application layer [9] which is made possible thanks to RDF.

2.2 Interoperability Strategies and Models

One of the first models that we relate to interoperability and has focus on being core for reuse is The Dublin Core. The Dublin Core is a metadata element set intended to facilitate discovery of online resources. Originally conceived for author-generated description of Web resources, it has also attracted the attention of formal resource description communities such as museums and libraries [10]. These elements and concepts have semantics representing the lowest common denominator for describing resources. As such, the Dublin Core is not intended to replace richer models, but delivers a core set of descriptions that can be used directly for resource description rather than serve as the basis for implementing more domain specific models.

In the UK, the strategy started from the data itself. Opening up the UK government data (data.gov.uk) emphasized why and how Linked Data was introduced and how a web of linked government data was created as part of the Linked Open Data cloud rather than focusing on making different datasets interoperable [11]. Therefore, there was no urge for reusable core data models. This strategy was preceded by a debate on selecting a closed or open warehousing model [12]. A tendency at the time of writing was that data storage causing a high demand for metadata integration, which is in current terms translated to the need for a convergence on semantics of that data storage and at the time already implied the need for standardization [13].

In the US, financials departments oblige all software to be compliant with the XBRL core data model. Semantics are expressed in the form of metadata within the XBRL taxonomy. XBRL provides a global standard for expressing business rules without relying on the application layer. Because of this, every stakeholder interacting in the information value chain can use and have a consistent understanding of the data's meaning [14]. Also in Europe, XBRL is being used for financial reporting between the financial sector and regulatory supervisors[5].

The Interoperability Solutions for European Public Administrations (ISA) Programme promotes interoperability across multiple interoperability levels (technical, semantic, organisational and legal) of Member State's borders and public service sectors. Its key components to ensure solutions are the European Interoperability Framework (EIF)[1], the European Interoperability Reference Architecture (EIRA) and the Common Assessment Method for Standards and Specifications (CAMSS).

To demonstrate the relation between core data models in general and interoperability, an example case (the ABC-model) on multimedia interoperability used a core model as the underlying framework and represented the applicable vocabularies as RDF schema class and property hierarchies. It demonstrated that it is possible to ascertain the intersections, differences, and domain-specific aspects of each of the underlying ontologies [15]. This has enabled to determine the most appropriate attachment points on the core data model, as an indicator for the interoperability.

There are many ways to achieve data interoperability which includes the conceptualisation of the high-level data integration workflow. Essentially recurring steps include schema alignment, data mapping, entity reconciliation, and data alignment. One of the examples implementing this workflow in a governmental context, discusses its application to a practical data integration exercise concerning Czech public procurement data in the field of business and organisation information [16].

3. DATA GATHERING

In this section we discuss how data was gathered. We started from the input of participants of the international Community of Practice on Core Data Models, a network of representatives from public administration organisations aiming to share knowledge, experience and lessons-learnt on core data models during a workshop in Brussels in November '14 [6]. Participants were expert representatives of organisations in the EU Member States that are working on core data models, either for their own purposes or in the context of cross-sector/-border standardization initiatives. Representatives from organisations from non-European countries, who are involved with the topic in the same organisational contexts also took part. In total 37 experts participated, representing organisations from 13 countries (11 EU Member States, Japan and the USA). Additional data was gathered via literature and document analysis.

3.1 Vocabulary Reuse

In the core data models, several common vocabularies are being reused or referred to. Public administrations can use and extend the Core Vocabularies (i) as a default starting point for designing the conceptual and logical data models in newly developed information systems. In the context of information exchange between systems, (ii) the Core Vocabularies can function as a basis for context-specific data model used to exchange data among existing information systems. For data integration (iii) they can be used to integrate data that comes from disparate data sources and create a data mesh-up. And for open data publishing (iv), the Core

[5]XBLR Europe. www.xblreurope.org. Accessed 2015-12

[6]Workshop Core data models for public administrations. Ibidem. Accessed 2015-12

Vocabularies can be used as the foundation of a common export format for data in base registries like land registries, business registers and service portals [17].

There are some relevant commonly used vocabularies focusing on a specific domains: to describe persons there is the *Friend of a Friend* (FOAF) vocabulary[7]. The *Simple Knowledge Organization Scheme*[8] (SKOS) represents classification schemes and *DCType*[9] (DC) and *DCTerms*[10] (DCTERMS) are commonly used for metadata types and terms are. *Semantically Interlinked Online Communities*[11] (SIOC) connects internet blogs, forums and mailing lists. *DCAT*[12] is a vocabulary designed to facilitate interoperability between data catalogs published on the Web. *VCARD*[13] is used to describe contact information and *Organization*[14] (ORG) is an ontology for organisational structures, aimed at supporting publishing of organisational information across a number of domains. *XML Schema*[15] (XSD) and *RDF Schema*[16] (RDFS), are also most often used to annotate data types or object types respectively.

As mentioned above, the following core data models have been developed so far under the ISA Programme [17]:

- **Core Person**: captures the fundamental characteristics of a person, e.g. the name, the gender, the date of birth, the location.
- **Registered Organisation**: captures the fundamental characteristics of a legal entity (e.g. its identifier, activities) which is created through a formal registration process, typically in a national or regional register.
- **Core Location**: captures the fundamental characteristics of a location, represented as an address, a geographic name or geometry.
- **Core Public Service**: captures the fundamental characteristics of a service offered by public administration.

The 'Handbook for using the Core Vocabularies' [18] indicates how public administrations can design domain data models and information exchange data models as extensions of the Core Vocabularies in a specific context and with a chosen syntax binding. The Core Vocabularies have a conceptual data model and several syntax bindings (UML, RDF, XML).

Existing Models

In this chapter we inventory the existing models. First we describe how they are mapped according to the core vocabularies. This allows us to filter on those models that are practical implementations.

Table 1 shows for each of the RDF syntax binding of the core vocabularies, the common Linked Data vocabularies each of them reuses and introduces.

3.2 Existing Models

In this chapter we inventory the existing models. First we describe how they are mapped according to the core vocabularies. This allows us to filter on those models that are practical implementations.

Table 1. Core data models and the common vocabularies each of them reuses

Core Data Models ISA	Vocabularies Introduced	Linked Data vocabularies properties and classes reused
Person[18]	person	dcterms; foaf; rdfs; schema; person
Location[19]	locn	dcterms; rdfs
Business[20] (W3C Regorg)	rov	dcterms; org; skos; xsd
Public Service[21]	cpsv	dcterms; foaf

The Danish **Grunddatamodellen** (Basic Data Model), is a composite, logical data model describing central public data domains. It is a UML class diagram, describing the consolidated public data to become available on the Data Distributor by the end of 2015[22].

The **IMI Core Vocabulary** developed in the Japanese IMI project provides a Concept Dictionary (description of meaning, relationship and hierarchical structure) and sample Information Exchange Package Descriptions for fundamental terms which can be accessed on the project website[23].

The **National Information Exchange Model** (NIEM) focuses on understanding, standardizing and ensuring data is discoverable across US public administrations. NIEM represents an approach to drive standardized connections among and between governmental entities as well as with the private sector and international partners which enable disparate systems to share, exchange, accept, and translate information [19]. The common data connections developed using NIEM result in reusable artifacts that reduce future

[7] Brickley, D. Miller, L. FOAF Vocabulary Specification 0.99: 2014. http://xmlns.com/foaf/spec/. Accessed 2015-12.

[8] W3C. 2011. SKOS. *http://www.w3.org/2004/02/skos/core#* Accessed 2015-12.

[9] Dublin Core: 2014. DCMI Metadata Terms. *http://purl.org/dc/dcmitype/*. Accessed 2015-12.

[10] Dublin Core: 2014. DCMI Metadata Terms. *http://purl.org/dc/terms/* Accessed 2015-12.

[11] SIOC Core Ontology Namespace: *http://rdfs.org/sioc/ns#*. Accessed 2015-12.

[12] Data Catalog Vocabulary: 2014. *http://www.w3.org/ns/dcat#*. Accessed 2015-12.

[13] VCARD. *http://www.w3.org/2006/vcard/ns#*. Accessed 2015-12.

[14] The Organization Ontology. *http://www.w3.org/ns/org#*. Accessed 2015-12.

[15] XMLSchema: 2014. *http://www.w3.org/2001/XMLSchema#* Accessed 2015-12.

[16] W3C. The RDF Schema vocabulary (RDFS). *http://www.w3.org/2000/01/rdf-schema*. Accessed: 2015-12.

[17] ISA. Improving semantic interoperability in European eGovernment systems. *http://ec.europa.eu/isa/actions/01-trusted-information-exchange/1-1action_en.htm*. Accessed: 2015-12.

[18] ISA Programma Person Core Vocabulary. *http://www.w3.org/ns/person*. Accessed: 2015-12.

[19] ISA Programme Location Core Vocabulary. *http://www.w3.org/ns/locn*. Accessed: 2015-12.

[20] W3C. Registered Organization Vocabulary (REGORG). *http://www.w3.org/ns/regorg*. Accessed: 2015-12.

[21] Grunddatamodellen. *http://data.gov.dk/model/model.htm*. Accessed: 2015-12.

[22] Grunddatamodellen. *Ibidem*. Accessed: 2015-12.

[23] Infrastructure for Multilayer Interoperability. *http://imi.ipa.go.jp*. Accessed: 2015-12.

development costs, covers the same goal as the XBRL standard discussed in section 2.2.

UBL, the Universal Business Language[24] is a library of standardized electronic XML business documents such as purchase orders and invoices. UBL is designed to plug directly into existing business, legal, auditing, and records management practices, eliminating the re-keying of data in existing fax- and paper-based supply chains and providing an entry point into electronic commerce for small and medium-sized businesses. The United Nations Centre for Trade Facilitation and Electronic Business, (**UN/CEFACT**)[25], trade and administrative organisations, from developed, developing and transitional economies, to exchange products and relevant services effectively - and so contribute to the growth of global commerce.

A cornerstone of the approach is the Core Component Technical Specification (**CCTS**)[26] which defines a meta model and rules necessary for describing the structure and contents of conceptual and logical data models and information exchange models and has a formalisation in UML.

The **OSLO** Vocabulary[27] is a simplified, reusable and extensible data model that captures the fundamental characteristics of information exchanged by Flemish (Belgian) public administrations in the domains of contact information, public services and localisation. It is developed by a multidisciplinary Working Group with experts from 28 organisations and representatives of the ISA programme. OSLO had the opportunity to adapt to the core vocabularies from early on in the development process. The goal was to support data interoperability from the beginning, which implied preparing the data infrastructures in advance with the right data models. This was guaranteed on the (core) ontology (semantic convergence) level and the data level. Support groups, mailing lists and open maintenance can help ensuring reuse of this ontology. In Flanders, a prototype for a distributed shared catalogue of public services and products from municipalities made explicit its applicability [20].

Stelselcatalogus, developed by the Dutch government, is an online catalogue of definitions of all concepts that are included in the Dutch base registers and related legislation. It defines concepts and metadata to get insight in the relations of the base registers. It facilitates reusers such as policy makers, IT'ers and administrative law experts to find out which source to use[28].

The **Swedish Company Data Model** (Bolagsverket)[29], instigated by the Swedish Companies Registration Office, has created conceptual data models (Grundläggande uppgifter om företag) of core concepts and information exchange objects that are used by Swedish public administrations to represent companies. These concepts and information exchange objects include for example Legal Person, Natural Person, Address, etc.

The German **XÖV** Kernkomponenten are generic core components, which serve as the basis for the creation of specific data models, created by the KoSIT, the German Coordination Cell for IT-standards[30].

3.2.1 Mappings

Earlier self-assessment exercises provided us information on the degree of overlap with the core vocabularies and each of the specific models as described in the Core Data Model Mapping Directory[31]. We plotted the summary of the models to come to a ranking. Firstly, per model we grouped the match degrees (narrow to exact) to compute the match percentage (degree of overlap) (see figure 2). Secondly, we grouped all the models to gain insights in the different degrees of overlap (see figure 3).

The SKOS classifiers[32], 'exact', 'broad', 'narrow', 'close', 'related', 'no', were used to indicate the overlap. We follow the definitions, as explained: (i) Two terms are an exact match if they describe exactly the same. They are equivalent. (ii) No match means two terms have nothing in common nor depict a related concept. They are completely disjoint. (iii) They have a close match if the set of subjects in one is mostly equal to the set of subjects of the other. (iv) There is a related match as soon as there is a meaningful intersection between the subjects of both. Mapping should be implemented with caution. (v) A narrow match occurs when the set of subjects in one is a superset of the set of subjects in the other. The definition of one generalizes the definition of the other. Finally, (vi) a broad match depicts the opposite of a narrow match.

Because the core data models are designed to be reused in other more applied contextual data models it is relevant to measure to which degree there is an overlap. Based on the self-assessment exercise and the SKOS classifiers, we explain the overlay (matches) with existing initiatives in relation to the core vocabularies below in figure 2. In this public mapping document, the core vocabulary concepts are assessed against particular vocabularies and classifications and are also mapped to the computed the degree of overlap.

---

[24] UBL v2.1. *http://docs.oasis-open.org/ubl/os-UBL-2.1/UBL-2.1.html*. Accessed: 2015-12.

[25] UN/CEFACT. *http://www.unece.org/cefact/*. Accessed: 2015-12.

[26] Core Components Technical Specification: 2009. *http://www.unece.org/fileadmin/DAM/cefact/codesfortrade/CCTS/CCTS-Version3.pdf*. Accessed: 2015-12.

[27] OSLO - Open Standards for Linked Administrations in Flanders. *http://purl.org/oslo*. Accessed: 2015-12

[28] Stelselcatalogus. *http://stelselcatalogus.nl*. Accessed: 2015-12.

[29] Grundläggande uppgifter om företag. *http://uppgiftskrav.bolagsverket.se/*. Accessed: 2015-12

[30] XÖV-BIBLIOTHEK: 2014. *http://www.xoev.de/de/bibliothek*. Accessed: 2015-12

[31] Core Data Model Mapping Directory: 2015. *http://mapping.semic.eu/*. Accessed: 2015-12

[32] SKOS Simple Knowledge Organization System: 2012. *http://www.w3.org/2004/02/skos/*. Accessed: 2015-12.

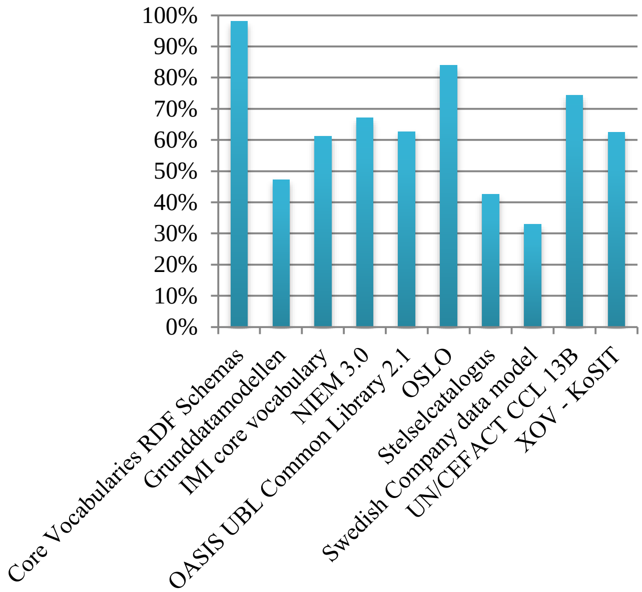

Figure 2. Match percentages for each of the core data model implementations

The relatively low scores of some implementations are shown in Figure 2, especially the Grunddatamodellen, Swedish Company Data Model and the Stelselcatalogus. Figure 2 shows as well that OSLO has the strongest overlay, due to the fact that OSLO was created as an extension of the Core Data Model and the awareness of the core data models was there from the start. We expect that over time as the awareness institutionalizes, other implementations will follow and the overlap and reuse of core data models will only increase. For some data models, like the Grunddatamodellen, the score is not representative because they are still in the development phase and at the moment only support concepts related to address and organisation.

The core data models are intrinsically intended as building blocks. They are not intended to be copied exactly but rather to support modular shaping of the data model implementations to eventually lead to better interoperability in the domain of application. Furthermore the chart in Figure 3 indicates that indeed, even when there are matches and overlap with the core data models, these matches are not always exact. They vary from broad over related to narrow matches.

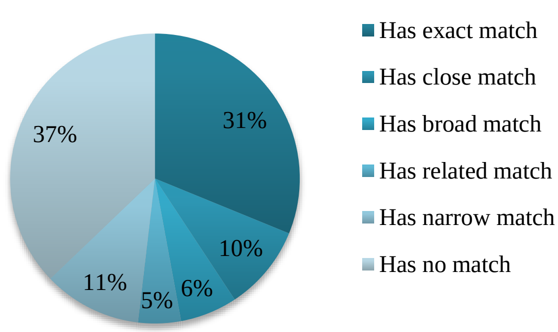

Figure 3. Matches with the core data models overall implementations

3.2.2 Selected Examples

Based on the overview of the core data models vocabularies and mappings, a selection was made to analyse the creation, use, maintenance and coordination. We focused on practical implementations of the core vocabularies and take parallel projects such as OASIS UBL and UN/CEFACT not into account.

Selected examples then include: Grunddatamodellen (DK), IMI Core Vocabulary (JP), NIEM 3.0 (US), OSLO (BE), Stelselcatalogus (NL), Swedish Company Data Model (SE) and XÖV Kernkomponentent (DE).

4. ANALYSIS CRITERIA

In the previous sections we selected core data models that will be evaluated. In this section, we introduce criteria to evaluate the current state of the art of core data models. These criteria refer to the lifecycle of core data models, as described in a detailed review published by the SEMIC action in late 2014. In this study, an analysis of a methodology and tools for the management of core data models and reference data for EU Institutions and Member States is described. [21]

Suarez-Figueroa et al identify in their Waterfall Ontology Network Life Cycle Model approach three pillars of ontology development: (i) a focus on the development process and the activities related to it, (ii) the life cycle referring to the process and order of the related activities and (iii) a focus on the methodology, describing methods on how to carry out the activities. More than an analysis of activities *as such* related to each data model, our research focus targets finding trends and communalities in terms of governance and (technical) management of core data models' lifecycle [22]. As there is no one-size-fits all approach to the lifecycle models of ontologies and data models we highlight in a simplified framework elements that will serve as evaluation criteria, as shown in figure 4. In the next paragraphs we operationalize these elements: the creation and development of each specific core data model (4.1), the use of each model (4.2) and the maintenance and coordination aspects (4.3).

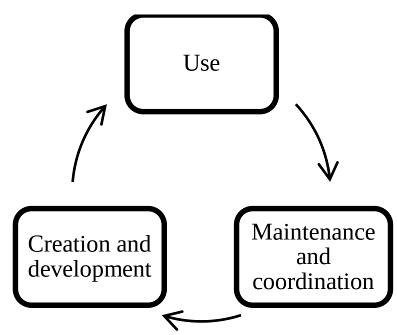

Figure 4: Evaluation criteria based on the lifecycle of core data models

4.1 Creation

This evaluation criterion is divided in four sub-criteria to clear out differences and similarities on how core data models are created.

- **Processes approach.** First, a distinction is made between top-down versus bottom-up approaches in terms of the development process. This means that a small group of experts prescribe a model on the basis of generic requirements. Bottom-up approaches involve combining existing data models and requirements from individual organisations.

- **Initiator.** Secondly, to evaluate the creation process, we identified the initiating organisations or actors that pushed off the creation process. This could be ministerial departments responsible for a specific policy domain, governmental agencies, third parties.
- **Timeline.** Thirdly, we bring the starting time of the creation of core data models into scope.
- **Decision mechanism.** This last sub-element describes how decisions on the data model are made. Within data modelling, this mostly happens via consensus building efforts. Different levels of participation are possible either within public administration or using an open consultation process.

4.2 Use

This global criterion discusses (i) the reasons to use core data models, (ii) the mechanisms to support the use of the models, and (iii) what the actual use of the models is.

- **Reasons and motivations for use.** When new data models are created, organisations can use core data models as a basic building brick that can be extended and designed for domain specific challenges. The use and reuse of core data models promotes as well the access to base registers. By creating mappings to core data models, existing data models are harmonized.
- **Mechanisms to support the use.** In providing core data models, concrete methodologies described in handbooks and manuals on the use must be provided. These can include rules for aspects of design and naming. Other systems that can support the promotion of the use of core data models are collaborative platforms, supporting co-creation and expert feedback. Licences define the appropriateness of (re-)using the core data models and rights on them.
- **Real use / Beyond proof-of-concept.** To what extent is the data model used or in which development phase is the core data model situated? When not, what are barriers prohibiting uptake?

4.3 Maintenance and Coordination

This criterion evaluates the level of maintenance and coordination of created and implemented core data models. In particular we evaluate the:

- **Approach for coordination and maintenance.** This criterion evaluates how the maintenance of the core data models take place and what the premises behind the approach are. Of interest are the elements of the maintenance process that are in use, such as instruments and processes to capture change requests, impact analysis and the application of real changes.
- **Mechanisms for (technical) coordination.** The scope of this criterion is to identify the practices used to capture changes related to the core data model that might occur. In particular, how the core data model implementation is documented plays an important role in its maintenance and coordination. We discern two important indicators:
  - *Formalisation***:** How the domain model is described in a technical specification outlining the entities, relations and attributes.
  - *Serialisation***:** The translation of the domain model to a *machine-interpretable* schema, typically RDF or XML Schema.

5. ANALYSIS

In this section, we evaluate the selected examples described in section 2.2 on the criteria explained in section 3 of this paper. We evaluate these models to come to descriptions of tendencies of the creation, use, coordination and maintenance of core data models.

5.1 Creation

The reasons to create core data models are fueled by the intentions to (i) promote the use and give access to base registers in a country (OSLO, Grunddatamodellen, Swedish Company data model, Stelselcatalogus) (ii) to avoid miscommunication between administrations, (iii) to harmonize data models (e.g. Core Vocabularies, XÖV) and (iv) to simplify cross-border exchange of information via the publication of data and via messaging (NIEM, IMI).

A base register is a trusted, authentic source of information under the control of an appointed public administration or organisation appointed by government. Base registries provide basic information on items such as persons, companies, vehicles, licences, buildings, locations and roads.

Base registers are the cornerstone of public services; they are closely related to Master Data in enterprises. This is the authoritative, most accurate data that is available about key business entities, used to establish the context for business transactions and transactional data.

We distinguish between top-down versus bottom-up approaches in the development process of core data models. Predominantly a top-down approach is taken for the development of core data models when a small group of experts prescribe a model on the basis of generic requirements (Grunddatamodell, Stelselcatalogus, Swedish Company data model). Bottom-up approaches involve combining existing data models and requirement from individual organisations (NIEM, OSLO). In the case of IMI, there's a hybrid approach, combining a top-down process for the creation of core data models, and a bottom-up process for the development of domain-specific data models.

We also evaluated which organisms initiate the creation of core data models. We find that in the case of Grunddatamodell, IMI, Stelselcatalogus, Swedish Company Data Model and XÖV, specific Ministries or (external) governmental agencies are the initiating organisms. The NIEM and OSLO data models were developed by intermediary non-governmental organisations active in the field of information exchange and e-government.

An evaluation of the decision mechanisms related to content, changes and structure by the creation of core data models, are mostly built on consensus (NIEM, OSLO, Swedish Company Data Model) or based on legal framework that define the possibilities (Grunddatamodel, Stelselcatalogus, IMI).

Can we detect a relation between the approach taken, the responsible organisms and the applied decision mechanisms? It seems likely that when a central governmental actor initiates the creation of core data models, the approach taken is mostly top-down and the decision mechanism is defined in legislation. In the opposite direction, we cannot state the argument that when the creation of a core data model is instigated through a bottom-up process that consensus building (or a hybrid form) is more applied as decision mechanism. This could be explained by the fact that building and working towards consensus through the stimulation of support and acceptance can stimulate future uptakes of particular core data models by other administrations. Another reason as well could be that the initiative for creating a core data model process starts from lower governmental levels, as central government is rather reluctant to it in the first phase (OSLO).

5.2 Use

During the workshop participants declared the reasons to use core data models and how these uses are steered and controlled in order to support its use.

We evaluate the use of core data models on three levels: firstly, we look at those core data models that are actually implemented and used in public administrations. Secondly, we investigate the main reasons for using the core data models. Thirdly, we analyze the mechanisms that support the use.

First, we note that the Grunddatamodellen, IMI and OSLO are still in the conceptual modus, while the other models are already implemented. Reasons not to use (barriers) are the fear of losing control (related to data quality, the lack of expertise of translating existing domain-specific data models into the core; resistance towards the 'new' and the lack of communication and promotion.

Second, the main reasons and motivations for use are highly related to the reasons to create core data models, as discussed in section 5.1. On the one hand, we find that striving at better *ex ante* alignments between administrations is one of the drivers to use core data models. This is the case where the core data model serves as a basic organisation-wide model for new data models and descriptions (Grunddatamodellen) or in a specific context, such as contact information (OSLO) and business reporting (Swedish Company Data Model). This is as well in the case of NIEM, where new data model's compatibility (mapping, creation of subsets, extensions) is in line with NIEM and extend afterwards its specialties. This is one of the advantages as the form of the compatibility is guaranteed with the use of RDF. On the other hand, reasons for use of the core data model aims at *ex post* alignment, where the model facilitates a better harmonization between existing data models. This is the case of XÖV.

Table 2. Evaluation of the creation of core data models

Core Data Model	Process	Initiator	Timing	Decision mechanism
Basic Data Programme Grunddatamodellen (DK)	Top-down	Ministry of Finance	2015	Defined in legislation
IMI Core Vocabulary (JP)	Hybrid	Information-technology Promotion Agency	2013	Consensus
NIEM	Bottom-Up	Global Justice Information Sharing Initiative	2005	Consensus
OSLO (BE)	Bottom-Up	External mediator representing local governments	2012	Consensus
Stelselcatalogus.nl (NL)	Top-Down	Ministry of Interior	2009; 2013 (v2)	Defined in legislation
Swedish Company Data Model (SW)	Top Down	External government Agency: e-Delegation	2013	Consensus building based on legislation
XÖV (GE)	Top Down	External government agency: KoSit	2008	Consensus building

Thirdly, we identify mechanisms to support the use of core data models: the creation of handbooks, user guides or directions towards administrations are a common practice, instigated by Grunddatamodellen, Stelselcatalogus, NIEM and XÖV. These latter two prescribe as well a specific license granting third parties the right to use the core data models and create derivative works. Interesting is the case of Stelselcatalogus, that is set up as the default practice to stimulate the use of the data model. Following the idea of 'Comply or explain', external administrations should proactively indicate why they will not be able to use the prescribed core data model. To stimulate future uptake of the OSLO model, the Flemish central administration invests in training for local administrations to put the OSLO specifications as a requirement in tender descriptions.

5.3 Maintenance and Coordination

In this section we evaluate the maintenance and coordination aspects of these core data models used in real implementations.

5.3.1 Coordination and Maintenance Approach

Ideally, coordination and maintenance of core data models starts from a holistic approach. This means that maintenance is aligned and coordinated for all produced derivatives, such as methodologies, tools, UML profiles, naming and design rules. This is considered a challenge, as each core data model was created and developed within a particular administrative and organisational context in different countries.

Table 3. Evaluation of the use of core data models

Core Data Model	Reasons and motivations for use	Mechanisms to support the use	Real use / Beyond proof-of-concept
Basic Data Programme – Grunddata-modellen (DK)	Basic model as an outset for new models and descriptions	(Design) Rules	Development phase Foreseen implementation in In Data Distributors Metadata Registry
IMI Core Vocabulary (JP)	Ensuring public information is well understood by the public and by internal eGov systems	User Guides	Behind trial stage; new implementation plans
NIEM (US)	Better information exchange, agreement on meanings in a broad range of fields.	Rules and User guide – Collaborative platform – specific license	High use in all US states; uptake in Canada and Australia; implemented release cycle
OSLO (BE)	Better information exchange and as an answer to local vendor-lock in	Tender specifications, training of civil servants	Proof-of-Concept
Stelselcatalogus. NL)	Consultation of base registers	'Comply or explain'	Partial uptake
Swedish Company Data Model (SW)	To be extended model for basic business reporting	Central real-time catalogue	Proof-of-concept in 1 key project
XÖV – Kernkomponenten (GE)	Harmonisation of existing data models	Specific licence; handbook; certification	300.000 msgs. a day are sent between municipalities; +25 specifications

By setting up a release cycle for model updates, changes can be incorporated on a predictable and sustainable schedule. In the case of NIEM, communication and alignment are installed between committees for technical architecture, business architecture and the NIEM-community. The NIEM Core stays the same until a new major release. Individual domain updates can happen any time. The coordination task is executed by the NIEM program. Unlike, in the XÖV core components are part of XÖV framework. There's no fundamental release cycle, but updates are possible mostly anytime and independently of one another, because of its focus on an overall open process of XÖV products. An external government agency is appointed by the Federal Government to operate this framework and the maintenance and coordination of XÖV. In the case of OSLO there is no formal maintenance and coordination program. Updates are occurring ad-hoc and often in case of new development projects making use of open standards.

5.3.2 Mechanisms for (Technical) Coordination

As the goal of core data models is to be reused to the maximum extent, stability is necessary which means that the models need to stay as fix as possible. Expansions (modular?) of the models are mostly allowed by adding new concepts, but not by changing existing concepts. This can be achieved by means of a separation between the conceptual, logical and physical data models (XÖV).

Another way to manage changes is to allow references to previous versions of the core model, which is the case with the NIEM implementation.

Stimulating a binding syntax is different for each core data model and is related to the degree of separation between the conceptual, logical and physical level. NIEM has a strict syntax binding because they do not foresee a separation on the conceptual, logical and physical level. Because of this separation XÖV has a less strict syntax binding. The coordinating office has developed a tool which visualizes the explanations of the various syntax bindings or how data model specifications link to the core data model. OSLO is not formally adopted by the relevant administrations, which implicates it is more vulnerable for the impact of changes.

5.3.3 Formalization and Serialization

This additional flexibility when linking, is facilitated by NIEM by following the principles of RDF. However important aspects of its implementation are translated in a dedicated XML schema. This introduces several limitations: there are no external reusable URI's and identifiers are only usable within the contents of a package (a single information exchange or implementation). In fact RDF offers more flexibility than an XML schema, because it has the important property of being modular: it does not enforce other implementations to strictly follow or inherit the schema. As RDF models graph structures: the union of two descriptions (as directed graphs) is mapped into the union of the corresponding RDF structures; this means that in presence of partial information the output is still a consistent RDF model, that can be successfully processed (thus without strictly adhering to a certain schema. *OSLO*, *IMI Core Vocabulary* and the *Stelselcatalogus* are examples of core data model implementations that use RDF as a data model and thus benefit from the additional flexibility in linking to other data models. Data models such as the *Grunddatamodellen* are only implemented on a conceptual level and serialized the schema and its properties in the XML Metadata Interchange (XMI) format. However, a conversion from XMI to RDF via Complete Meta-Object Facility (CMOF) is planned. The *Swedish Company Data Model* and *XÖV* both offer a core component conceptual model in the Unified Modelling Language (UML) with accompanying specification.

In particular for the latter of the two, it means that core concepts are maintained on the conceptual level, syntax is to be added in a later stage. This is different from *NIEM*, *OSLO* or the *IMI core vocabulary* that aims to provide reusable specification and serializations. This has the benefit making changes easier to the model, but specific information exchange specification and design rules need to be explain with sufficient amount of detail nonetheless. Immediate reuse is not directly promoted but it clearly encourages use by providing a clear how-to.

6. DISCUSSION AND FUTURE WORK

We started this paper from the lifecycle of core data models, referring to the evolution and growth of the models and represent the relations between creation, use and coordination. The main reasons of existence of core data model is better data exchange and interoperability between administrations.

We find relations between creation, use and coordination, particular in the approach that the models find their implementation: top down versus bottom-up.when ex ante alignment of data models between administration occur, rule systems (e.g. coercion via legislation) and descriptions seem to be necessary to make the core data model work. When ex post creation and use, consensus building is more in place. A top down perspective could be more effective, e.g. when implemented and supported it in to central legislation (as seen in the PSI-directive by the European Parliament). A bottom-up approach is therefore more creative, starting from (ICT) structures at the basis administrations often with a strong engagement. On the other hand, these initiatives are often more chaotic and is more time-consuming to come to a joint-vision.

This brings us to suggestions for two directions of further research from a governance point of view. A first direction is on the coordination and governance level of core data models. A critical question raises on the feasibility of the idea of "core data models" that harmonise data model design across domains and across sectors. There is an increasing need for it, as e.g. European programs are developed in this area towards its member states, but it is not sure if involved governments, companies and standard bodies can attain the required level of coordination. Hence, coordination and control within eGovernment projects are often hard to determine, because participating administrations all can have their own perspective on the goal and on the intergovernmental relations. Creating insight and vision on how to instigate communities of practice and how to institutionalise and coordinate these ways of collaboration, especially in a cross-country and even in a cross-continental context, is needed. These coordination matters (as well from a technical point of view) are necessary in order to obtain interoperability in a cross-country and even in a cross-continental context, especially on the conceptual and logical level. The choices for certain syntax bindings (that occurs as a best choice in a specific practice then serves) then should be applied as supporting and enhancing core data model implementations.

On a more high end level, we suggest to compare information systems and the way European governments (and in this administrations) are structured. In particular we focus on the aspect of autonomy, coercion and the concept of federalism. Both in public administration research, political science and in information management studies, the concept of federations (or federated systems) occur (eg as described by authors as Burgess [23] and Breton [24]. For example, the European Union is seen as an example of centripetal federalism, where nation states relinquish parts of their autonomy to the higher policy level. Centrifugal federalism occurs then when forces attract autonomy from the central to the lower institutional level. We suggest to investigate in depth the level of occurrence, the development and the coordination actions between these in relation to core data models.

7. CONCLUSIONS

In this paper we made an inventory of on the implementations of international Core Data models for public administrations. These reusable data models are defined, created and maintained to facilitate interoperability across different systems, applications and domains, as public administrations are often still organized in vertical silos. Their main goal is to enhance a better information exchange between administrations, within nations and beyond. We evaluated the Grunddatamodellen (DK), IMI Core Vocabulary (JP), NIEM 3.0 (US), OSLO (BE), Stelselcatalogus (NL), Swedish Company Data Model (SE) and XÖV Kernkomponentent (DE) based on the criteria that refer to the life cycle of core data model. In particular we invested the similarities, differences and tendencies on the (i) creation and development, (ii) use and (ii) maintenance and coordination of the models. These insights are the basis to suggest further directions for researchers in the field of policy studies and information management.

8. ACKNOWLEDGMENTS

The research activities described in this paper were funded by Ghent University, the Institute for the Promotion of Innovation by Science and Technology in Flanders (IWT), the Fund for Scientific Research-Flanders (FWO-Flanders), and the Interoperability Solutions for European Public Administrations (ISA) Programme of the European Commission.

9. REFERENCES

[1] ISA. 2010. *European Interoperability Framework (EIF) for European public services*. p. 9. http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf.

[2] European Commission. 2014. *Study on eGovernment and the Reduction of Administrative Burden*. p. 149.

[3] Pardo, T.A., Nam, T. and Burke, G.B. 2011. E-Government Interoperability: Interaction of Policy, Management, and Technology Dimensions. *Social Science Computer Review*. p. 10.

[4] Colpaert, P., et al. 2014. Quantifying the Interoperability of Open Government Datasets. *Computer*, 2014(10): p. 50-56.

[5] Bovalis, K., et al., *Promoting Interoperability in Europe's E-Government.* Computer, 47(10): p. 25-33.

[6] Wood, D. 2011. Linking government data. *Springer Science & Business Media.*112 p.

[7] Janowicz, K., et al. 2014. Five stars of linked data vocabulary use. *Semantic Web*, 5(3): p. 173-176.

[8] Colpaert, P., et al. 2014. Painless URI dereferencing using The DataTank, in *The Semantic Web: ESWC 2014 Satellite Events*, Springer. p. 304-309.

[9] Melnik, S. and Decker, S. 2000. A layered approach to information modeling and interoperability on the web. In *Proc. of the ECDL'00 Workshop on the Semantic Web*. (Lisbon, Portugal, september 18-20, 2000).

[10] Weibel, S. 1997. The Dublin Core: a simple content description model for electronic resources. *Bulletin of the American Society for Information Science and Technology*, 24(1): p. 9-11.

[11] Sheridan, J. and Tennison, J. 2010. Linking UK Government Data. In *LDOW 2010*, (North-Carolina, April 27, 2010). p.4.

[12] Vetterli, T., Vaduva, A. and Staudt, M. 2000. Metadata standards for data warehousing: open information model vs. common warehouse metadata. *ACM Sigmod Record*, 29(3): p. 68-75.

[13] Chan, L.M. and Zeng, M.L. 2006. Metadata interoperability and standardization-A study of methodology, Part II. *D-Lib magazine*, 12(6): p. 4.

[14] Debreceny, R., et al. 2010. Does it add up? Early evidence on the data quality of XBRL filings to the SEC. *Journal of Accounting and Public Policy*, 29(3): p. 296-306.

[15] Hunter, J. 2003. Enhancing the semantic interoperability of multimedia through a core ontology. *Circuits and Systems for Video Technology, IEEE Transactions on*, 13(1): p. 49-58.

[16] Mynarz, J. 2014. Integration of public procurement data using linked data. *Journal of Systems Integration*, 5(4): p. 19-31.

[17] ISA. *CORE VOCABULARIES*. [Action description] 2014; Available http://ec.europa.eu/isa/actions/documents/eudig12a-1401-i01-core-vocabularies-lr_en.pdf.

[18] European Commission. 2014. *Handbook for using the Core Vocabularies.* ISA Programme. p.10. https://joinup.ec.europa.eu/site/core_vocabularies/Core_Vocabularies_user_handbook/Handbook-for-using-the-Core-Vocabularies_v0.50.pdf.

[19] Federal Counsel. 2010. *Agency Information, Exchange Functional, Standards Evaluation. Adoption and use of the National Information Exchange Model (NIEM).* USA Federal CIO Counsel. p. 120.

[20] De Vocht, L., et al. 2014. Converging on semantics to ensure local government data reuse. In *Proceedings of the 5th workshop on Semantics for Smarter Cities (SSC14), 13th International Semantic Web Conference (ISWC)*. 6 p.

[21] European Commission. 2014. *Metadata management requirements and existing solutions in EU Institutions and Member States*. ISA Programme. http://ec.europa.eu/isa/documents/metadata-management-requirements-and-existing-solutions-in-eu-institutions-and-member-states_en.pdf.

[22] Suarez-Figueroa et al. 2012. *Ontology Engineering in a Networked World*, Springer-Verlag Berlin Heidelberg: p. 25-27.

[23] Burgess, M. 2006. *Comparative federalism: Theory and practice.* Routledge. 349 p.

[24] Breton, A. 2000. Federalism and Decentralization: Ownership Rights and the Superiority of Federalism. *The Journal of Federalism* 30 (2). 1