Modeling Digital Preservation Capabilities in Enterprise Architecture

Christoph Becker, Gonçalo Antunes, José Barateiro, Ricardo Vieira, José Borbinha
INESC-ID Information Systems Group, Lisbon, Portugal
becker@ifs.tuwien.ac.at, {goncalo.antunes,jose.barateiro,rjcv,jlb}@ist.utl.pt

ABSTRACT

The rising awareness of the challenges of preserving information over the long term has led to a wealth of initiatives developing economic models, methods, tools, systems, guidelines and standards for digital preservation. The challenge of digital preservation is to assure that information nowadays coded and stored in digital formats can be read and be used in an unforeseen future. This is an interdisciplinary problem combining organizational and technical challenges. However, to date there is no unified view on how to approach the problem from a holistic perspective and align organizational and technical issues in a systems engineering approach. Organizations that aim to *add* digital preservation to their abilities generally have difficulties to assess their existing systems and what capabilities and components they are missing in order to address the needs of trustworthy information longevity.

In this paper we present an approach that enables us to accommodate the concerns of digital preservation in Enterprise Architecture practice. We discuss key elements of a generic reference architecture for digital preservation and a capability model based on established domain-specific reference models. Distilling these knowledge sources into a consistent and coherent view allows baseline assessment and incremental capability development in typical IT governance scenarios where an IT architecture already exists. We illustrate this with the assessment of a government agency's existing capabilities and systems against emerging digital preservation requirements.

Categories and Subject Descriptors

H.1 [**Information Systems**]: Models and Principles; J.1 Administrative Data Processing Government; K.6.4 Management of computing and Information Systems

General Terms

Management, Documentation, Design, Standardization

Dg.o'11 June 12-15, 2011, College Park, MD, USA
Copyright 2011 ACM 978-1-4503-0762-8/11/06 ...$10.00.

Keywords

Enterprise Architecture, Digital Preservation, IT Governance, Standards

1. INTRODUCTION

In a digital information age where ubiquitous, transparent access to trustworthy information is at the heart of civil society, government agencies have readily understood their mission to secure the knowledge and information of their society over the long-term [28].

Digital Preservation (DP) is a core area where IT problems and solutions intersect with organizational policies and missions, where long-term visions are to be implemented by IT solutions that are inherently short-lived and ever-changing. Digital material is constantly threatened on the physical level (the data carriers), the logical level (the representation of information in data structures and file formats), and the semantic level (the meaning of content presented to a human). Robust technical solutions are available to address the physical storage level; international research efforts are strongly focusing on the logical level of keeping content accessible over changing environments. However, the problem also needs to be addressed from the information system's perspective. From this perspective, it is at its core a problem of aligning IT and business – an intersection of IT research and the problems and missions of an enterprise. This is the core mission of Enterprise Architectures, which have in the last decades been driven strongly by the defense domain [7], but also achieved wide acceptance in industry [27].

Initiatives on DP have been pushed largely by cultural heritage institutions [28]. The Reference Model for an Open Archival Information System (OAIS) [15] provides a high-level view of an archival organization and has been very influential in the DP domain. However, it is by far not the only reference to consider when embarking on a DP effort. Criteria catalogs for trustworthy repositories specify requirements a repository should fulfill to be trustworthy [6, 9]. These criteria are spanning all levels of organizational and technical responsibilities and are often very hard to evaluate. Similarly, the Records Management community has developed criteria and models for supporting archives in their quest to secure authenticity and provenance of their holdings [14, 4].

All these references can be seen as *domain knowledge bases* distilling the partial knowledge of a certain community. They are of tremendous value in enabling communities to communicate and compare their approaches. However, these knowledge bases are not without internal inconsistencies, and are not necessarily aligned with each other. Moreover, many of the criteria and guidelines in these documents overlap with well-established problem areas such as Information Security and Risk Management, for which a solid body of knowledge already exists that should not be neglected.

In recent years, numerous national and international institutions have started initiatives to build or acquire a Trustworthy Digital Repository (TDR). A recent survey showed that '... many organizations are beginning to make a transition from analyzing the problem to solving it. They remain concerned that mature solutions do not yet exist. Nevertheless, 85 percent of organizations with a digital preservation policy expect to make an investment to create a digital preservation system within two years. Such systems are likely to be componentized, mix-and-match solutions.' [25] Procurement of these systems is still notoriously difficult without a clear understanding of the alignment of existing system components, capabilities and processes with the specific processes and capabilities required by a TDR. The OAIS provides only a very high-level and narrow view on the principal functions of such a system. Moreover, it prescribes a certain solution architecture that does not necessarily fit in an organization's IT landscape. In describing an almost monolithic, separated *system* for DP, it complicates, if not precludes, the concept of incrementally adding capabilities and components to an existing system, such as an Enterprise Content Management System, to enable it for DP.

In this paper, we address this gap between heterogeneous, potentially conflicting reference models and domain knowledge bases. We present a coherent architectural approach that enables interdisciplinary business-IT alignment and supports communication between problem owners and solution providers by accommodating the concerns of DP in Enterprise Architecture practice. We present a consistent and coherent *architecture vision* combining established Enterprise Architecture frameworks with domain-specific knowledge bases and best-practice models. This allows an organization to express its drivers and constraints, main goals, key performance indicators, stakeholders, and desired capabilities, in a homogeneous and modular way and thus to assess its current *baseline architecture* for its ability to provide the required capabilities in a consistent manner. It can thus provide a bridge to enable interdisciplinary business-IT alignment between agencies responsible for long-term archival of content and solution providers offering (parts of) the systems required to achieve this long-term mission.

This paper is structured as follows. Section 2 outlines related approaches and standards in the areas of DP, IT Governance, and Enterprise Architectures. Section 3 presents key aspects of accommodating these domain concerns within TOGAF's Architectural Development Method and outlines key architectural elements. Section 4 illustrates the relation between goals, capabilities and services in a real case of a municipal archive that is currently undergoing a controlled change process to add DP capabilities to its operations. Finally, Section 5 draws conclusions and gives an outlook on current and future work.

2. RELATED WORK

Three major areas converge in the presented work: Trust and Digital Preservation; IT Governance as a more general perspective; and Enterprise Architecture as the framework that helps to unify viewpoints and align business and IT. We will discuss each area in turn to give a rough overview of the body of knowledge it contributes to the overall picture.

2.1 Trust and Digital Preservation

The mission of digital preservation (DP) is "to ensure continued access to digital materials... it refers to all of the actions required to maintain access to digital materials beyond the limits of media failure or technological change" [19]. The discourse in DP has always been interdisciplinary and dominated by a need for standardization and common language. Current efforts in DP build upon the OAIS model to address the fundamental issues surrounding trust and provide a certification standard for digital repositories [6, 17].

In the Records Management field, the central question of authenticity in digital records has been of equal importance [10], leading to early standardization efforts [14]. The current revision of the *Model Requirements for the Management of Electronic Records* (MoReq2010) [8] is a substantial catalog of functional requirements for an electronic record management system (ERMS) that covers aspects ranging from classification schemes through audit trails, backup, recovery and security to referencing, searching, and retrieval. It is much more grounded in formal modeling than the OAIS, but with its hundreds of requirement statements, it is often overwhelming in size and complexity. Moreover, it covers not just the core DP capability of an ERMS, but its entire functionality, and delivers minute detail on the desired *operation* of specific components of an ERMS. It also comes with a metadata model based on a number of existing standards.

Several other standards define vocabularies and schemas for storing and exchanging metadata as well. PREMIS is probably the most widely used scheme [23], but as a recent visualization illustrated, the list is immense[1]. On a more strategic level, a recent initiative analyzed the economic side of the preservation problem and provided recommendations for catalyzing sustainable developments for DP [3].

DP is essentially information management with a long-term perspective. Information management thus covers records management and archives, but includes any kind of collection, management, and distribution of information. This in turn requires strong information *technology* management and IT governance.

2.2 IT Governance

IT Governance is a key discipline for making effective decisions and communicating the results within IT-supported organizations. Its main purpose is to identify potential managerial and technical problems before they occur, so that actions can be taken to reduce or eliminate the likelihood and/or impact of these problems. *Control Objectives for Information and related Technology* (COBIT) [5] is a set of best practices, measures and processes to assist the management of IT systems. COBIT is not specific to a technological infrastructure nor business area, and intends to fill the gap between requirements, technical issues and risks. It includes a framework, a set of control goals, audit maps, tools to support its implementation and, especially, a guide for IT management. The latter is organized in the domains of (*i*) Planning and Organization; (*ii*) Acquisitions and Implementation; (*iii*) Delivery and Support; and (*iv*) Monitoring and Evaluation. These processes address the areas of strategic alignment (alignment of IT with the business) [11]; value

[1] http://www.dlib.indiana.edu/~jenlrile/metadatamap/

Acronym	Description	See
OAIS	The Open Archival Information Systems Model (OAIS, ISO 14721) is a high-level reference model for an archival organisation with a long-term responsibility for providing understandable information to a specified user community.	[15]
ISO 20652	The CCSDS Producer-Archive Interface specifies interactions between content producers and an OAIS archive.	[16]
TDR	Trusted Digital Repositories: Attributes and Responsibilities was an early milestone in specifying criteria for trustworthiness in digital repositories.	[24]
PREMIS	The PREMIS (Preservation Metadata: Implementation Strategies) Data Dictionary provides a vocabulary of entities relevant in digital preservation.	[23]
TRAC	Trustworthy Repository Audit and Certification Criteria and Checklist (TRAC) is the basis for an ongoing certification standardisation initiative and widely referenced throughout the cultural heritage and data preservation communities.	[6, 17]
PP	Preservation planning is a core element of DP, comprising the evaluation and decision making necessary to ensure the right actions are taken to keep information understandable.	[2]
SHAMAN RA	The SHAMAN Reference Architecture presents a generic view of a digital preservation architecture	[1]
MoReq2010	The Model Requirements for the Management of Electronic Records (MoReq) draft 0.92 specifies functional requirements for an Electronic Records Management System.	[8]
Zachman Framework	The Zachman Framework provides a holistic projection of the levels and dimensions of an enterprise architecture that supports the reconciliation of viewpoints and the assessment of coverage of domain-specific models in terms of the enterprise's key aspects	[29]
TOGAF ADM	The Open Group Architecture Framework (TOGAF) is a leading EA framework comprising a set of methods, tools, content models and guidelines to support architecture development. At its core is the TOGAF Architecture Development Method (ADM).	[27]
BMM	The Object Management Group's Business Motivation Model provides a framework for defining ends and means of an organisation and thus can fulfill an important role in reconciling existing policy frameworks in DP.	[22]
COBIT	The Control Objectives for Information and related Technology (COBIT) is a key IT management standard.	[5]
IEEE 1540	The IEEE 1540-2001 Standard for Software Life Cycle Processes - Risk Management addresses risks within the software lifecycle process.	[13]
ISO 27000 series	The ISO/IEC 27000 series include a set of requirements, code of practice, implementation guidance, evaluation metrics and a risk managment process to establish and Information Security Management System.	[18]

Table 1: Key sources used in digital preservation architecture development

delivery (creation of business value); resource management (proper management of IT resources); risk management; and performance management.

The "ISO/IEC 27000 series" [18] include a set of standards developed for information security matters. This family of standards specifies the Information Security Management Systems (ISMS) Requirements, proposing a process approach to design, implement, operate, monitor, review, maintain and improve an ISMS. The *design* process follows a risk management approach, including the definition of the risk assessment approach, risk identification, risk analysis, evaluation of risk treatment options and selection of controls to treat risks. The requirements proposed in these standards intend to be generic and applicable to all types of organizations, independent of type, size and nature.

2.3 Enterprise Architecture

From the discussion above, it becomes clear that even the tip of the iceberg already provides an impressive array of standards on all levels and a wealth of models that should be considered. Table 1 provides a short and necessarily incomplete summary of some of the core references we have been actively relying on in the present work. What is clearly needed is a unified view and a practical way of approaching the problem in the real world. Here it is that Enterprise Architecture comes into play.

Architectural descriptions provide rigorous descriptions of complex systems with diverse concerns, and are a recommended approach to tackling the dynamic and increasing complexity of those systems. According to the IEEE Std. 1471-2000, which has also become ISO/IEC 42010:2007, architecture is "the fundamental organization of a system, embodied in its components, their relationships to each other and the environment, and the principles governing its design and evolution" [12]. It considers that a *system* has a *mission* and inhabits an *environment* which influences it. It also has one or more *stakeholders* that have *concerns* regarding the system and its mission. Concerns are "those interests that pertain to the system's development, its operation, or any other aspects that are critical or otherwise important to one or more stakeholders".

A system has an *architecture* described by an *architecture description* which includes a rationale for the architecture. The architecture description is also related with the stakeholders of the system and deals with several *views* according to the *viewpoints* of the stakeholder. This includes functional and non-functional aspects of stakeholders' concerns.

Accurate architecture descriptions provide a "complete picture" of the overall system. However, any system (especially a complex system made of software, people, technology, data and processes) is continuously subject to changes, usually driven by the evolution of the system environment [21].

Enterprise Architecture then is a holistic approach to systems architecture with the purposes of modeling the role of information systems and technology in the organization, aligning enterprise-wide concepts and information systems with business processes and information. It supports planning for sustainable change and provides self-awareness to the organization [26]. In that sense, it aims to provide a complete coverage of the organization.

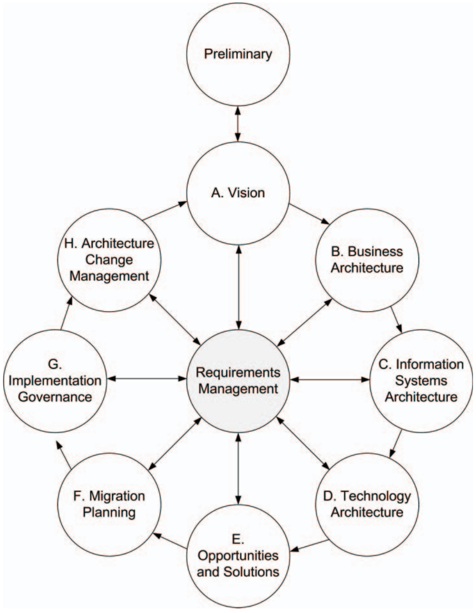

Preliminary
A. Vision
B. Business Architecture
C. Information Systems Architecture
D. Technology Architecture
E. Opportunities and Solutions
F. Migration Planning
G. Implementation Governance
H. Architecture Change Management
Requirements Management

Figure 1: TOGAF Architecture Development Method (ADM)

The Zachman framework is a ”way of defining an enterprise’s systems architecture” with the purpose of ”giving a holistic view of the enterprise which is being modeled” [29]. It is also presented as a ”classification theory about the nature of an enterprise” and the kinds of entities that exist within[2]. The Zachman framework presents itself as a table where each cell can be related to the set of models, principles, services and standards needed to address the concerns of a specific stakeholder. The rows depict different viewpoints of the organization (Scope, Business Model, System Model, Technology Model, Components, and Instances), and the columns express different perspectives on each of the viewpoints (Data, Function, Network, People, Time, Motivation). Due to this visually appealing nature, it is very useful in analyzing the scope of specific models and frameworks, and in reconciling potentially conflicting viewpoints.

The Open Group Architecture Framework (TOGAF) [27] provides methods and tools to support architecture development. It comprises seven modules which can be partly used independently of each other. The core of TOGAF are the Architecture Development Method (ADM) and the Architecture Content Framework. The ADM as the heart of TOGAF consists of a cyclical process divided in nine phases as shown in Figure 1. After a *preliminary* phase in which the context, relevant guidelines and standards as well as the goals of the architecture process are identified, the main process begins with the elaboration of an *architecture vision* and the principles that should guide the architecture. This architecture vision provides the basis for developing the business architecture, information systems architecture, and technology architecture. On this basis, solutions are developed, and migration and implementation are planned and governed. Finally, Architecture Change Management ensures that the architecture continues to be fit for purpose. The ADM can be adapted for various purposes, and in more complex situations, the architecture can be scoped and partitioned so that several architectures can be developed and later integrated using an instance of the ADM to develop each one of them.

Similar to TOGAF in its claim, the Department of Defense Architecture Framework (DoDAF) supports the unification of the architectures of different military commands, services and agencies to improve decision making and information sharing [7]. DoDAF describes a six-step process for the development of architecture descriptions and a series of non-prescriptive viewpoints which can be used on the architecture development process. These viewpoints are composed of models and supported by the DoDAF Metamodel, which guides the architecture content and formally defines the vocabulary for architecture development. In particular, the *capability viewpoint* supports incremental acquisition and visualization of evolving capabilities to support procurement in complex situations.

2.4 Observations

A number of models, standards and criteria catalogs may guide, constrain and ultimately confuse decision makers on the quest of a digital preservation *capability.* The fields that are destined to have a say in creating a solution to a certain DP problem range from IT Security and Risk Management to metadata standards as well as overlapping and potentially conflicting compliance criteria from specific domains.

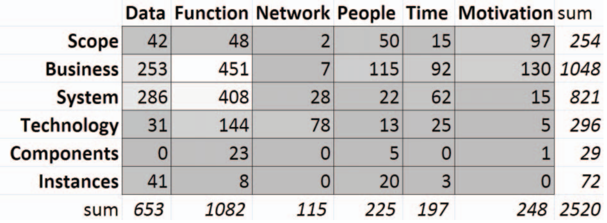
Most of the initiatives in the broad DP field have been developing models to facilitate communication *within* a closed community. Using these models for communication with related communities encountering the very same *core DP* problem of keeping information understandable over time, but in different organizational environments, has been notoriously difficult. Even harder is it to convey the knowledge contained in these models to a general IT procurement, management and operation audience. National and international cooperation is a key to success for digital preservation, but preservation partnerships have been hindered substantially by ‘competing priorities, lack of funding, lack of knowledge, and different perspective of IT people’ [20]. The Reference Architecture presented in [1] presents an important step towards a more holistic view on the digital preservation problem, based on Enterprise Architecture concepts. However, it is not based on an in-depth analysis of existing domain knowledge bases to a degree that enables their convergence in a transparent manner.

In the next sections, we will discuss an Enterprise Architecture approach based on TOGAF and discuss some of the issues arising in the application of the TOGAF ADM to combine the sources outlined here. We will further present the key concepts of an *architecture vision* that results from this process.

3. ADDRESSING DP CONCERNS IN ENTERPRISE ARCHITECTURE

A large number of reference models, standards, catalogs of requirements and criteria for trustworthiness on various levels, and other source materials exist. A systematic approach is needed to reconcile and align concepts from the main references in the domain, enable reuse of well-defined concepts and best-practices and thus improve common understand-

[2] `http://www.zachmaninternational.us/index.php/ea-articles/100-the-zachman-framework-evolution`



	Data	Function	Network	People	Time	Motivation	sum
Scope	42	48	2	50	15	97	254
Business	253	451	7	115	92	130	1048
System	286	408	28	22	62	15	821
Technology	31	144	78	13	25	5	296
Components	0	23	0	5	0	1	29
Instances	41	8	0	20	3	0	72
sum	653	1082	115	225	197	248	2520

Figure 2: Mapping of TRAC in the Zachman Framework

ing of the domain by modeling its knowledge in a reusable fashion. Such a *reference architecture* constitutes a process from which multiple architecture artifacts can result. We are using TOGAF's ADM to develop a capability-centered reference architecture which addresses DP concerns and considers the main references in the DP domain as key inputs. We thus discuss the accommodation of DP concerns in the two key phases that ADM commences with: *Preliminary* and *Architecture Vision*.

3.1 Preliminary: Reconciliation of sources

The *Preliminary* phase of the ADM consists of the preparation and initiation of the architectural activities and includes the definition of the principles that will govern the architecture work. During this phase, the internal and external organizations impacted by the architecture work are assessed and defined, and the key reference models established. Subsequently, key references are analyzed and merged in order to create a general understanding of the domain. That understanding creates the conditions to initiate the Architecture Vision phase in which key concepts for a DP architecture emerge. We will in this section discuss critical aspects encountered in the reconciliation and alignment of potentially conflicting reference models. In the next section, we discuss key concepts of the architecture vision.

Domain-specific 'knowledge bases' or 'reference models' may contain

- Elaboration of typical stakeholder concerns, actors and their goals and interests;
- Mandatory requirements, i.e. constraints, necessitated by external influencers such as legal situations or other non-negotiable requirements commonly encountered;
- Contracts and governance metrics;
- Domain concepts and corresponding design patterns for domain models, roles and interactions;
- Design patterns and building blocks for solutions; and
- Value propositions for functions and systems with or without reference to actors or stakeholders.

This implies that these aspects need to be considered at different stages of the architecture development cycle. Furthermore, domain references may contain statements that can be interpreted in different ways and span different architecture concerns. Merging disparate sources is only feasible based on a clear distinction between these categories of statements and a clear definition of terms.

We have relied on two techniques to facilitate contextualization and alignment of knowledge bases: (1) The Zachman Framework provides a basic grid of alignment onto which statements can be mapped. (2) Concept maps are both a practical tool for visualizing concepts and a formal graph model that can be queried and statistically analyzed.

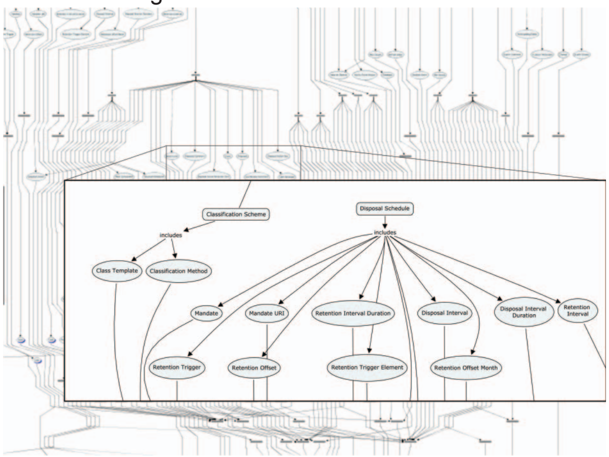


Figure 3: A MoReq2010 concept map

3.1.1 *Basic analysis using the Zachman Framework*

Using the Zachman Framework as a projection space, we can map statements piece by piece onto its cells to develop an understanding of the concerns that a source covers. For instance, TRAC consists of 84 statements of the kind '*B3.2 Repository has mechanisms in place for monitoring and notification when representation information (including formats) approaches obsolescence or is no longer viable.*' with associate explanation and examples. Terms such as 'mechanism' are left undefined and thus open for interpretation. We conducted a group exercise where every participant for every statement got 10 points to distribute across the cells of the Zachman framework. Summing these scores over participants, one can obtain a common understanding of the maximum coverage of concerns of single statements, groups, and the totality of statements. Figure 2 displays a visualization of the overall result of such an exercise with 3 participants for the complete TRAC document. While on this level we do not get a detailed nor exact view on specific statements, it is clearly visible that some aspects are *not* considered by TRAC, while the bulk of statements concern functions on data on the business and system level.

3.1.2 *Concept consolidation using Concept Maps and graph analysis*

TRAC is a list of criteria created from a business viewpoint, but often concerning very low-level technical aspects and solution components. This makes it hard to represent it as constraints and requirements and thus make it operational. Most of the statements in TRAC cover multiple cells in the Zachman grid, not necessarily constrained to one column and one row. In fact, some span areas as distinct as Motivation/Scope and Data/Instances! This lack of separation of concerns means that a translation and decomposition into the distinct, but related, aspects of these criteria is required, while keeping full traceability to the original source and its intent. This can be achieved by representing key concepts in a graph structure such as concept maps.

We created a representation of both TRAC and MoReq2010 as concept maps using CMapTools[3], resulting in 14 maps for TRAC's 84 criteria and 17 for MoReq2010 (which contains 792 requirement statements). Figure 3 shows part of a general concept map of MoReq2010 key concepts, magni-

[3] `http://cmap.ihmc.us/`

Concept *relation* Concept	TRAC Source
Repository *participates in* Deposit Agreement Negotiation	A5.1, A5.2, A5.3, A5.5
Producer/ Depositor *participates in* Deposit Agreement Negotiation	A5.1, A5.2, A5.3, A5.5
Deposit Agreement Negotiation *results in* Deposit Agreement	A5.1
Depositors and Other Relevant Parties *participates in* Negotiation	A5.3
Repository *participates in* Negotiation	A5.1
Negotiation *results in* Contracts and Deposit Agreements	A5.1, A5.2, A5.3
Service Level Agreements *negotiated with* Producers	A3.1
Repository *negotiates* pre-accessioning agreements	B1.1
Producer/ Depositor *negotiates* pre-accessioning agreements	B1.1

Table 2: Concept relationships containing 'negotiat'

fying some concepts associated with retention schedules and classification. Clearly, such a map can become very complex. While the direct analysis of this map is hardly feasible, concept maps are graphs and can thus be queried to allow statistical analysis. Annotations can be used to provide full traceability to (potentially multiple) root sources a concept has originated from. Table 2 shows results of a full-text query for 'negotiat%' on the combined TRAC concept maps, providing full requirements traceability to the source statements. We notice some redundancy about 'deposit agreements' which occur in different flavors. Interestingly, the top ten relations in the TRAC concept maps are *has* (84 occurrences), *such as* (40), *has in place* (10), *is* (9), *documents* (8), *contains* (7), *defines* (6), *identified in* (6), *is one of* (6), and *consists of* (6). By further expressing concept relationships such as equivalence, specialisation and enumeration, it becomes possible to align disjunct groups of statements.

3.2 An Architecture Vision

The *Architecture Vision* phase includes the definition of scope, the identification of stakeholders and their concerns and the elaboration of a value chain; constraints, drivers, goals, and key performance indicators; and finally, capabilities and the envisioned solution architecture.

3.2.1 Stakeholders and Concerns

The identification of the main stakeholders of digital preservation and their concerns used the main references of the DP domain, resulting in thirteen different stakeholders. The IEEE Std. 1471-2000 [12] considers that stakeholder identification must take into account, at minimum: (i) the users of the system; (ii) those responsible by the acquisition and governance of the system; (iii) the developers and providers of the system's technology; and (iv) the maintainers of the system as a technical operational entity.

End-User related stakeholders include the *Producer/ Depositor*, the *Consumer*, and the *Designated Community.* The *Producer/Depositor* is the entity responsible for the ingestion of the objects to be preserved. It may be the owner of the object, but can also be any other entity entitled to perform this action. The terms Producer and Depositor are used interchangeably in distinct sources to describe the stakeholder responsible for content ingestion. The *Consumer* stakeholder represents the user accessing to the preserved objects, who has a potential interest in its reuse and a certain background in terms of knowledge and technical environment. The *Designated Community* is defined in OAIS as 'an identified group of potential *Consumers* who should be able to understand a particular set of information' [15]. This group can be characterized not only by domain knowledge, but also by technical means that are available to it, preferred usage scenarios, etc.

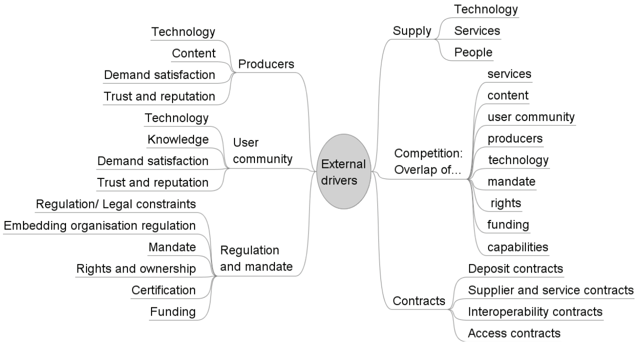


Figure 4: Mind map of the main classes of external drivers for DP

Manager stakeholders include the *Executive Management*, the *Repository Manager*, the *Technology Manager*, and the *Operational Manager. Executive Management* is responsible for strategic decision making on an organization level, ensuring that the mandate is fulfilled and the repository continues to serve its designated community. The Repository Manager is concerned with ensuring repository business continuity, defining business strategies and thus setting goals and objectives. That means it defines ends to be achieved by the repository and operates on the business domain, interacting with the designated communities, legal environment and constraints, etc. The *Technology Manager* is responsible for technological system continuity and the deployment of technological means to achieve the ends set by the repository business. The *Operational Manager* is concerned with the continuous policy-compliant operation of the repository, which involves balancing ends and means and resolving conflicts between them, i.e. constraints as set from *Technology Management* and *Repository Management.*

Compliance-related stakeholders include the *Regulator* and the *Auditor.* A *Regulator* is an external entity imposing rules concerning the preservation of digital assets, such as legislation and standards. These can apply to the organization or the system's technology and usage. The *Auditor* is responsible for certifying that the organization practices, the system's properties and the operational environments comply with established standards, rules and regulations.

Stakeholders concerned with operations include the *Repository Operator* and the *Technology Operator.* The *Repository Operator* is a business worker who may be aware of the details of the design and deployment of the system, but is primarily concerned with business, with no concerns about infrastructure management or strategic alignment. The *Technology Operator* is responsible for the regular operation and maintenance of the components of the technical infrastructure (hardware and software) and their interoperability, according to specified service levels.

Stakeholders concerned with solutions include the *System Architect* and the *Solution Provider.* The *System Architect* is responsible for the design and update of the architecture

Goal	Description	Example Key Performance Indicator (KPI)
G1	Acquire content from producers in accordance to the mandate, following agreed rules	*Number of distinct objects received per year*
G2	Deliver authentic, complete, usable and understandable objects to designated user community	*Percentage of transformational object properties preserved by actions as denoted by user feedback and/or QA measures in comparison to guarantees provided by specified SLAs*
G3	Faithfully preserve provenance of all objects and deliver accurate provenance information to the users upon request	*Percentage of access requests where objects' provenance is reported to be undefined, not clearly defined or wrong (e.g. indicated by number of incidents of fake objects reported)*
G4	Authentically preserve objects for the specified time horizon, keeping their integrity and protecting them from threats	*Percentage of legitimate access requests fulfilled successfully as denoted by user feedback*
G5	React to changes in the environment timely in order to keep objects accessible and understandable	*Average reaction time responding to obsolescence incident report*
G6	Ensure repository sustainability: mandate, technical, financial, operational, communities	*Time horizon of secured mandate greater or equal to average time horizon of objects*
G7	Build trust in the depositors, the designated community and other stakeholders	*Time horizon of mandate secured by legal means in years*
G8	Maximize efficiency in all operations	*Average yearly costs per object*

Table 3: Goals and Example KPIs

of the system and its alignment with the business objectives. The *Solution Provider* is concerned with providing components of the architecture. This may include software components, platforms and business services.

The specification of each stakeholder identified should include a concise elaboration of the concerns and key questions of interest to this stakeholder. For example, the *Producer/ Depositor* specification includes a concern *Acquisition of Content*: 'Submission of the objects along with the additional required data, using a supported interface, so that its preservation can be guaranteed according to the negotiated submission agreements and contracts.' Similarly, the Consumer is elaborated by two concerns: (1) '*Access*: Mechanisms are in place to ensure that contents are easily accessible using a supported interface according to agreements' and (2) '*Content*: The information retrieved is authentic, understandable and corresponds to my needs.' These concerns are associated with questions such as 'Will the objects be corresponding to my queries, authentic, compatible to my technical environment, and understandable?'

3.2.2 Drivers and Constraints

The definition of the stakeholder and subsequent analysis of their concerns enables us to define the general drivers and constraints of DP. A *driver* is "an external or internal condition that motivates the organization to define its goals", while a *constraint* is an "external factor that prevents an organization from pursuing particular approaches to meet its goals" [27].

The top categories of *internal drivers* are *Business Vision*, which is itself the highest level driver of the organization, and *Resources.* The former can be further divided in *Infrastructure*, *Hardware*, and *Software* and respective *Operational Costs*, existing *Capabilities*, and *Expertise* needed to ensure operations. *Staff* is also part of *Resource*-related drivers with associated *Personnel Costs*, existing *Expertise and Qualifications*, and *Commitment.*

External drivers are shown in Figure 4. The top categories are *Producers*, *User Community*, *Contracts*, *Supply*, *Competition*, and *Regulation and Mandate. Producers* can act as *external drivers* through the *Technology* they use, the *Content* they produce, the *Satisfaction of their demands*, and the *Trust and Reputation* the repository build or has built within the *Producer* community. The *User Community* can be a preservation *driver* for a organization through the *Technology* they use, their *Knowledge Base*, the *Satisfaction of their demands*, and the *Trust and Reputation* that the repository builds or has built within the *User Community. Contracts* can act as drivers, such as *Deposit Contracts*, *Supplier and Service Contracts*, *Interoperability Contracts* with other repositories, and *Access Contracts. Supply* can also act as a driver, namely at the level of desired and/or appropriate *Technology*, *Services*, and *Staff. Competition* is essentially defined by *overlaps* with other organizations that may require differentiation of the repository services. Finally, *Regulation and Mandate* presents itself as a key driver comprising a number of influences. *Regulation/Legal Constraints*, internal *Regulation* imposed by the organization responsible for the repository, the *Mandate* to preserve certain contents, *Rights and Ownership* concerning the objects to be preserved, the possible existence of *Certification* and a corresponding motivation to be certified, as well as the existence of sufficient *Funding*, are all examples of *Regulation and Mandate* drivers.

The definition of generic DP driver categories further allows the derivation of constraints. The category *rights and ownership*, for instance, may contain a constraint on the creation of derivative copies. Considering the driver category *Funding*, a constraint may be posed by funding that is insufficient to conduct quality assurance on preservation actions. The technology available to the user community, on the other hand, will constrain the choice of possible access channels.

3.2.3 Goals and Capabilities

The definition of stakeholders, drivers and constraints provides the basis for the specification of high-level goals for DP. Table 3 lists identified key goals and example Key Performance Indicators (KPIs). The first four goals represent the primary value chain, Goals 5 to 7 are required for successful goal achievement in changing environments, and Goal 8 adddresses efficiency. Quantifying the degree of fulfilment of these goals is an essential instrument for improvement: KPIs can only be used to assess the performance of an organization towards its goals when they are specific and measurable. This is sometimes hard to achieve and one of the key questions in DP. Any quantitative assessment of Goal 2, for example, requires an in-depth assessment of preservation

	Capability	Goal
Business	**Acquire Content**: The ability to offer services for transferring content from producers into the repository. This includes services for reaching agreement with producers about the terms and conditions of transfer. Realized by (the component capabilities) *Ingest* and *Ingest Negotiation*.	G1
	Secure Bitstreams: The ability to secure bitstreams for a specified amount of time (Bitstream preservation). Realized by *Bitstream Security Planning* and *Secure Storage Operation*.	G3, G4
	Preserve Content: The ability to maintain content authentic and understandable to the defined user community over time and assure its provenance (Logical preservation). Realized by *Preservation Planning* and *Preservation Operation*.	G3, G4, G5
	Disseminate Content: The ability to offer services for transferring content from the repository to the user community. This includes services for reaching agreement with users about the terms and conditions of transfer. Realized by *Discovery*, *Access*, and *Dissemination Negotiation*	G2
Support	**Data Management**: The ability to manage and deliver data management services, i.e. to collect, verify, organize, store and retrieve data (including metadata) needed to support the preservation business according to relevant standards. Realized by *Data Administration*, *Metadata Operations*, *Data Statistics and Reporting* and *Data Operations*	G2, G3
	Manage Infrastructure: The ability to ensure continuous availability and operation of the physical, hardware, and software assets necessary to support the repository.	G5, G6, G8
	Manage HR: The ability to continuously maintain staff which is sufficient, qualified and committed to performing the tasks required by the repository.	G6, G8
	Manage Finances: The ability to plan, control and steer financial plans and operations of the repository to ensure business continuity and sustainability.	G6, G8
Governance	**Manage Risks**: The ability to manage and control strategic and operational risks and opportunities to ensure efficient business continuity and sustainability.	G6, G8
	Compliance: The ability to verify the compliance of operations and report deviations.	G6, G7
	Community Relations: The ability to engage with the designated community and ensure that its needs are fulfilled	G5, G7
	Certification: The ability to obtain and maintain certification status	G6, G7
	Mandate Negotiation: The ability to negotiate mandates with governing institutions	G6, G7
	Business Continuity: The ability to identify business capabilities and assure mission-critical operations.	G5, G6
	Succession Planning: The ability to negotiate formal succession plans.	G6,G7
	IT Governance: The ability to manage and develop the services, processes and technology solutions that realise and support the primary capabilities.	G5,G6, G8

Table 4: Capabilities of concern in a Digital Preservation scenario

actions [2].

To achieve these goals, an organization will require certain capabilities. A *capability* is 'an ability that an organization, person, or system possesses' [27]. It is expressed in general high-level terms of its outcome and is not a business function, but a concept realized by a combination of elements such as actors, business functions and business processes, and technology.

Capabilities required for DP can be divided into business capabilities, governance capabilities, and support capabilities. While business capabilities concern the primary business goals and the value chain of an enterprise, support capabilities represent the ability to ensure continuous availability and operation of the infrastructure necessary to support the organization, including physical assets, hardware, and software. Governance capabilities enable strategic management of scope, context, continuity and compliance of the business. Table 4 describes the top-level capabilities that will generally be required by most organizations with the mission to preserve information and provide access to it for a specified group of consumers. The top-level business capabilities primarily address the core business goals; however, the capability to *preserve content* also has to be concerned with the fundamental goal to address change in the environment. The capability to preserve content is seen as vey distinct from the preservation of bitstream; this corresponds to the separation into bitstream and logical preservation. Support and governance capabilities address the remaining goals. These top-level capabilities are further decomposed into component capabilities.

This capability model supports partitioning of concerns and strives to make explicit the boundaries between distinct capabilities. Capabilities such as *Manage Infrastructure* are not a core DP concern, but the primary business capabilities in DP may pose specific requirements on the properties of the infrastructure that will have to be provided by the corresponding capabilities. On the other hand, content management systems provide discovery and dissemination services that are required by core capabilities in the DP model.

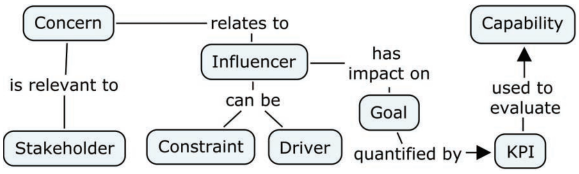


Figure 5: Main elements of the architecture vision

Figure 5 relates the main concepts of the architecture vision. A full specification of capabilities on all levels is beyond the scope of this article. Instead, we will focus on a specific case study and illustrate the chain of relations between specific drivers, goals and capabilities in a real-world case.

4. CAPABILITY ASSESSMENT CASE

The municipality of Lisbon (CML) is currently developing an infrastructure using the software Documentum[4] to support a wide set of well-defined business workflows for a wide range of organizational unities, including the Municipal Archives that are responsible for the record keeping and the archiving of the business information. Considering CML is a public body, a large number of this business information might be classified of historical public interest, so its dispos-

[4] http://www.emc.com/domains/documentum/index.htm

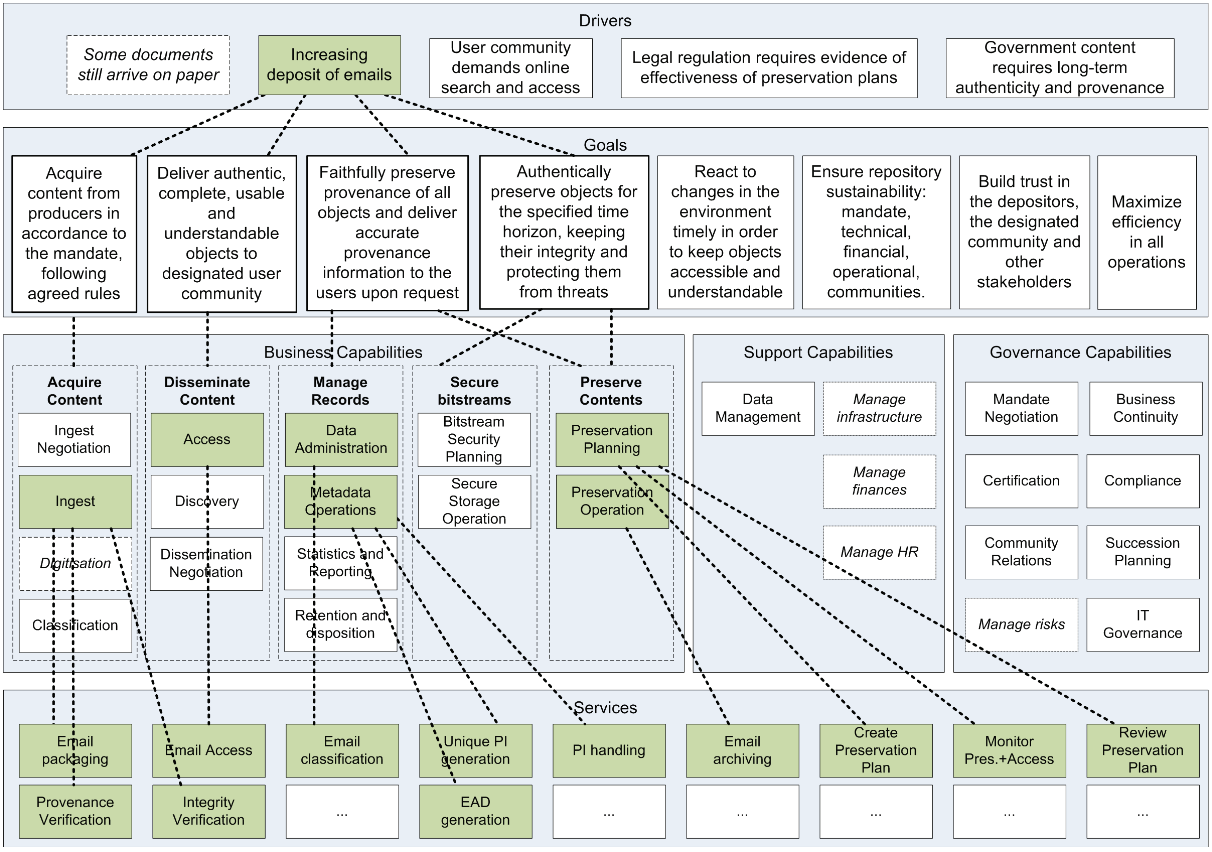

Drivers
Some documents still arrive on paper
Increasing deposit of emails
User community demands online search and access
Legal regulation requires evidence of effectiveness of preservation plans
Government content requires long-term authenticity and provenance
Goals
Acquire content from producers in accordance to the mandate, following agreed rules
Deliver authentic, complete, usable and understandable objects to designated user community
Faithfully preserve provenance of all objects and deliver accurate provenance information to the users upon request
Authentically preserve objects for the specified time horizon, keeping their integrity and protecting them from threats
React to changes in the environment timely in order to keep objects accessible and understandable
Ensure repository sustainability: mandate, technical, financial, operational, communities.
Build trust in the depositors, the designated community and other stakeholders
Maximize efficiency in all operations
Business Capabilities
Acquire Content
Ingest Negotiation
Ingest
Digitisation
Classification
Disseminate Content
Access
Discovery
Dissemination Negotiation
Manage Records
Data Administration
Metadata Operations
Statistics and Reporting
Retention and disposition
Secure bitstreams
Bitstream Security Planning
Secure Storage Operation
Preserve Contents
Preservation Planning
Preservation Operation
Support Capabilities
Data Management
Manage infrastructure
Manage finances
Manage HR
Governance Capabilities
Mandate Negotiation
Business Continuity
Certification
Compliance
Community Relations
Succession Planning
Manage risks
IT Governance
Services
Email packaging
Email Access
Email classification
Unique PI generation
PI handling
Email archiving
Create Preservation Plan
Monitor Pres.+Access
Review Preservation Plan
Provenance Verification
Integrity Verification
...
EAD generation
...
...
...
...
...

Figure 6: Part of a capability model for a municipal archive

able period can be undefined, which brings the concern of digital preservation to the front line.

In order to align existing services and capabilities with drivers and goals of the organization, we created a capability model linking drivers, goals, and capabilities to business services. Figure 6 presents a partial view of this model, with certain aspects highlighted for illustration. The five key drivers for the architecture effort shown on top are strongly originating from the content, producers' and users' technology, and legal regulations. These key drivers can be linked to the key goals of the archive, which are based on the generic goals described in Table 3. The influence can then be traced to top-level capabilities and further down to the affected services. We have highlighted one key driver for illustration, the increasing deposit of emails.[5] Capabilities and services affected by this driver are colored correspondingly. It can be seen that addressing the *increasing deposit of emails* requires a variety of services, ranging from integrity verification of emails during ingest to the creation of preservation plans [2]. Aspects relevant in this case, but considered outside the core scope of DP concerns, are given in *italics.* One example is presented by the driver *Some documents still arrive on paper*, which requires physical records management capabilities and a digitization capability as part of content acquisition. Furthermore, infrastructure and risk management capabilities are often a concern of repository planning and constitute a significant component of TRAC, but in fact represent core concerns of the respective areas of IT governance. It is worth noting that the goal of *building trust* is not necessarily a primary goal in such a scenario, since the depositors do not normally have a choice of which archive to rely on.

The capability assessment that is conducted on the basis of this model can serve not only as a bridge to enable communication and improve stakeholder involvement, but more generally as an entry point to a Reference Architecture approach and an enabler to incremental capability improvement. It can serve as a framework for formulating desired properties of capabilities and expressing criteria catalog that are generally accepted within one domain (such as TRAC and MoReq2010) as constraints on capabilities.

5. CONCLUSION

This paper discussed the problem of accommodating the concerns of digital preservation in Enterprise Architecture practice. We analysed issues arising in the reconciliation of potentially conflicting domain-specific knowledge sources, and we discussed tools that can be used to align viewpoints and foster common understanding. We presented key artifacts of a DP architecture, which include stakeholders and their concerns; drivers and constraints; goals and KPIs; and capabilities required to achieve these goals. The described capability model can be used to assess an organization's baseline architecture according to required DP capabilities.

A capability represents a manageable unit of change and supports incremental development through an explicit dis-

[5] Some terms in the diagram may require additional clarification. *Negotation* refers to any interaction with actors outside the system boundary to agree on terms of services, while *Acquisition* of content refers to the process of transferring content information into the archive as part of the Ingest. *EAD* means Encoded Archival Description.

tinction between systems and their capabilities. By developing a capability model concerned with digital preservation, it thus becomes possible to distill the essential requirements of DP, model the hierarchy of required capabilities and their dependencies, and transfer these capabilities from the originating *repository scenario* to any scenario where DP is of concern within a business context – i.e., where digital preservation is not core business, but a required *support capability.*

This Enterprise Architecture approach provides a coherent and consistent unified high-level view on specialized viewpoints in order to help control the complexity and inconsistency of information systems architecture with DP concerns through the adoption of architectural patterns and separation of concerns. It further improves strategic alignment through establishing the boundaries between DP concerns and other business concerns. This improves the definition of the problem and the assessment of the current and desired preservation capabilities.

Based on the work presented here, we intend to represent criteria catalogs such as TRAC and MoReq2010 as constraints on a Reference Architecture metamodel, elaborate the capability model, and develop a concise and well-documented reference architecture traceable to the original knowledge sources.

6. ACKNOWLEDGMENTS

This work was supported by FCT (INESC-ID multi-annual funding) through the PIDDAC Program funds and by the project SHAMAN, funded under FP7 of the EU under contract 216736.

7. REFERENCES

[1] G. Antunes, J. Barateiro, and J. Borbinha. A reference architecture for digital preservation. In *Proc. iPRES2010*, Vienna, Austria, 2010.

[2] C. Becker, H. Kulovits, M. Guttenbrunner, S. Strodl, A. Rauber, and H. Hofman. Systematic planning for digital preservation: Evaluating potential strategies and building preservation plans. *Int. Journal on Digital Libraries (IJDL)*, December 2009.

[3] Blue Ribbon Task Force on Sustainable Digital Preservation and Access. *Sustainable Economics for a Digital Planet.* 2010.

[4] CECA-CEE-CEEA, Bruxelles- Luxembourg. *Model Requirements for the Management of Electronic Records. MoReq2 Specification*, 2008.

[5] IT governance institute. CobiT 4.1. framework – control objectives – management guidelinces – maturity models, 2007.

[6] CRL and OCLC. Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC). Technical report, The Center for Research Libraries and Online Computer Library Center, February 2007.

[7] Department of Defense, Washington D.C. *DoD Architecture Framework, Version 2.0*, 2009.

[8] DLM Forum Foundation. *MoReq2010 - Model Requirements for Records Systems. Draft - v0.92*, 2010.

[9] S. Dobratz, A. Schoger, and S. Strathmann. The nestor catalogue of criteria for trusted digital repository evaluation and certification. In *Proc. JCDL2006*, 2006.

[10] L. Duranti. The long-term preservation of accurate and authentic digital data: The interpares project. *Data Science Journal*, 4(25):106–118, October 2005.

[11] W. V. Grembergen. *Strategies for information technology governance.* Idea Group Publishing, 2003.

[12] IEEE. *Recommended Practice for Architecture Description of Software-Intensive Systems (IEEE 1471-2000).* IEEE Computer Society, 2000.

[13] IEEE 1540-2001 - IEEE Standard for Software Life Cycle Processes - Risk Management, 2001.

[14] ISO. *Information and documentation: Records management (ISO 15489-1:2001)*, 2001.

[15] ISO. *Open archival information system – Reference model (ISO 14721:2003)*, 2003.

[16] ISO. *Space data and information transfer systems – Producer-archive interface – Methodology abstract standard (ISO 20652:2006)*, 2006.

[17] ISO. *Space data and information transfer systems - Audit and certification of trustworthy digital repositories (ISO/DIS 16363). Standard in development*, 2010.

[18] ISO/IEC 27001:2005. Information technology – Security techniques – Information security management systems – Requirements, 2005.

[19] M. Jones and N. Beagrie. *Preservation Management of Digital Materials: A Handbook.* Digital Preservation Coalition, London, UK, November 2008.

[20] H. Kwon, T. A. Pardo, and G. B. Burke. Building a state government digital preservation community: lessons on interorganizational collaboration. In *Proc. dg.o'06*, pages 277–284, NY, USA, 2006. ACM.

[21] T. Mens, J. Magee, and B. Rumpe. Evolving software architecture descriptions of critical systems. *Computer*, 43:42–48, 2010.

[22] Object Management Group. *Business Motivation Model 1.1.* OMG, May 2010.

[23] PREMIS Editorial Committee. *PREMIS Data Dictionary for Preservation Metadata version 2.1*, January 2011.

[24] RLG/OCLC Working Group on Digital Archive Attributes. *Trusted Digital Repositories: Attributes and Responsibilities.* Research Libraries Group, 2002.

[25] P. Sinclair, C. Billenness, J. Duckworth, A. Farquhar, J. Humphreys, L. Jardine, A. Keen, and R. Sharpe. Are you ready? Assessing whether organisations are prepared for digital preservation. In *Proc. iPRES2009*, 2009.

[26] P. Sousa, A. Caetano, A. Vasconcelos, C. Pereira, and J. Tribolet. Enterprise architecture modeling with the unified modeling language. In P. Rittgen, editor, *Enterprise Modeling and Computing with UML.* IGI Global, 2006.

[27] The Open Group. *TOGAF Version 9.* Van Haren Publishing, 2009.

[28] C. Webb. *Guidelines for the Preservation of Digital Heritage.* Information Society Division United Nations Educational, Scientific and Cultural Organization (UNESCO) – National Library of Australia, 2005.

[29] J. Zachman. A framework for information systems architecture. *IBM Systems Journal*, 12(6):276–292, 1987.