A Capability Model for Digital Preservation

Analyzing Concerns, Drivers, Constraints, Capabilities and Maturities

Christoph Becker
Vienna University of Technology
Vienna, Austria
becker@ifs.tuwien.ac.at

Gonçalo Antunes, José Barateiro, Ricardo Vieira
INESC-ID Information Systems Group, Lisbon, Portugal
{goncalo.antunes,jose.barateiro,rjcv}@ist.utl.pt

ABSTRACT

The last decade has seen a number of reference models and compliance criteria for Digital Preservation (DP) emerging. However, there is a lack of coherence and integration with standards and frameworks in related fields such as Information Systems; Governance, Risk and Compliance (GRC); and Organizational Engineering. DP needs to take a holistic viewpoint to acommodate the concerns of information longevity in the increasingly diverse scenarios in which DP needs to be addressed. In addition to compliance criteria, maturity models are needed to support focused assessment and targeted process improvement efforts in organizations. To enable this holistic perspective, this article discusses the question of capability maturity and presents a capability model for DP. We further demonstrate how such an architectural approach can be used as a basis to analyze the impact of criteria and metrics from the ISO Repository Audit and Certification standard on stakeholders, concerns, drivers, goals, and capabilities. The analysis presented here shall contribute to advance the understanding of cross-cutting concerns and the dicussion on maturity models in DP.

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

Keywords

OAIS Model, Repository Audit and Certification, Trust, Digital Preservation, Reference Architecture, Standards

1. INTRODUCTION

The last decade has seen considerable progress in clarifying the boundaries, goals and reference frameworks of DP. However, the relationships with related key disciplines such as Information Systems and Information Technology Management are still unclear. DP was originally driven strongly by the cultural heritage sector. Yet today, it is relevant for organizations in increasingly diverse business domains, ranging from the pharmaceutical sector to eScience and potentially any domain where information plays a key role. DP in an information-centric scenario is a cross-cutting capability orthogonal to the value chain. It has been increasingly found of fundamental importance for enabling the actual value delivery of organizations outside the traditional memory sector. DP operations are support functions to organizations that manage information and often intersect with information, services and technology across entire enterprises.

In the domain of DP, reference models for archival systems and corresponding compliance criteria have been developed. However, the general perspectives of fields such as Enterprise Architecture, Information Systems, and Governance, Risk and Compliance have not yet been fully considered. This poses a substantial barrier to increasing the recognition of DP in the mainstream fields of Information Systems and Information Technology. Furthermore, it has the effect that research in DP is often neglecting the conceptual models and powerful design techniques in fields such as Organizational Engineering, Software Engineering, and Information Systems.

The ISO 16363 standard is refining compliance criteria for repositories based on the OAIS Reference Model. The risk assessment method DRAMBORA[1] provides a catalogue of typical risks in DP environments [22]. These standards were developed specifically for traditional DP scenarios. Their focus on providing a system to address the DP problem as a whole makes it difficult to apply them in non-traditional DP settings. They deliver some guidance on compliance criteria to be met, but do not provide effective mechanisms for governance and control, or clear guidance on how to improve the processes of an organization with particular consideration of DP concerns. However, DP is becoming increasingly a concern in non-traditional environments, where the organizational environment may not be well suited for employing a DP *system* such as an OAIS-based approach, but instead requires an incorporation of DP abilities into the organizational and technological system, alongside existing processes and capabilities.

In this paper, we present a *capability model* for digital preservation that is based on established architectural principles and frameworks. We analyze this capability model

[1] `http://www.repositoryaudit.eu/`

from two perspectives. First, we discuss a *capability maturity model* based on CMMI and a method of assessing capability maturity for operational preservation. Second, we discuss the impact that criteria for trustworthy repositories as defined in ISO 16363 have on specific capabilities. The analysis presented shall be contributing to a clarification of maturity models in the field as well as an improved understanding in the implications that regulatory constraints, business drivers, and organizational goals have on organizational processes in the domain of DP.

This paper is structured as follows. Section 2 outlines related approaches and standards in the areas of DP, GRC, and Enterprise Architecture. Section 3 presents a capability model relating stakeholders and their concerns to drivers and constraints, goals, and capabilities. Section 4 discusses a maturity model for preservation operations. Section 5 relates the capability model to criteria for trustworthy repositories and illustrates the possibilities for analysis in organizational environments on a case study. Finally, Section 6 draws conclusions and gives an outlook on current and future work.

2. RELATED WORK

Digital preservation is a problem with many facets. It essentially surfaces in any organization that has to manage information over time. However, initiatives on digital preservation have been strongly driven by memory institutions and the cultural heritage sector [31]. The OAIS Reference Model [16] describes an information model and a conceptual model of key functional entities. It includes a high-level contextual view of an archival organization and its key stakeholders, and has provided a common language for the domain. However, it is difficult to reconcile these views with scenarios where different systems are in place, where related concerns may overlap with DP concerns and processes. This may for example occur in organizations where an Electronic Records Management System or an Enterprise Content Management System is in place. Key models in Records Management are the "Model Requirements for Records Systems' (MoReq2010) [12] and ISO 15489 [17]. Moreq2010 specifies functional requirements for an Electronic Records Management System and covers wide spectrum of aspects in hundreds of requirements statements. The Preservation Metadata Implementation Strategies (PREMIS) working group maintains a data dictionary for DP that contains intellectual entities, objects, rights, events, and agents [26] in a technically neutral model.

The 'Trusted Digital Repositories: Attributes and Responsibilities' report [27] (TDR) was a key milestone towards the standardization of criteria catalogs for trustworthy repositories. With the goal of providing audit and certification facilities, the Repositories Audit and Certification Criteria (RAC) are currently undergoing ISO standardization. They describe criteria for trustworthiness in the areas of Organizational Infrastructure; Digital Object Management; and Technologies, Technical Infrastructure, and Security [10, 19].

While these reference models deliver some guidance on compliance criteria to be met, they do not describe effective mechanisms for governance and control nor guidelines on implementation and improvement. However, they describe typical stakeholders and their goals and interests; recurring regulatory drivers and constraints; contractual structures, roles, and interaction patterns; solution practices and building blocks; and value propositions. As such, they are invaluable sources of domain knowledge.

DP problems, systems, and organizational concerns require a holistic, integrated view that combines aspects of organizational processes, contextual concerns, regulatory compliance and IT with systemic approaches for governance and control. These viewpoints are a stronghold of Enterprise Architecture (EA). The discipline of EA models the role of information systems and technology on organizations in a system architecture approach [15] in order to align enterprise-wide concepts, business processes and information with information technology and information systems. The core driver is planning for change and providing self-awareness to the organization in a holistic way [29]. The Zachman framework is a very influential early EA approach [32]. It describes the elements of an enterprise's systems architecture in a table where each cell is related to the set of models, principles, services and standards needed to address a specific concern of a specific stakeholder. The leading EA frameworks today are The Open Group Architecture Framework (TOGAF) [29] and the Department of Defense Architecture Framework (DODAF) [11].

IT Governance focuses on "the leadership, organisational structures and processes that ensure that the enterprise's IT sustains and extends the organisation's strategies and objectives" [8]. A widely known framework is COBIT: Control Objectives for IT. It provides a thoroughly defined process model linking resources, activities, processes and goals. One of the core concepts in Governance and Process Improvement is the idea of process *maturity*. It has been demonstrated that formal maturity models such as the Capability Maturity Model Integration (CMMI) are powerful tools for targeted improvement of processes based on quantitative assessment [14]. COBIT states that "... maturity modeling enables gaps in capabilities to be identified and demonstrated to management. Action plans can then be developed to bring these processes up to the desired capability target level" [8]. These target levels are defined in correspondence to the Software Engineering Institute's CMMI [7, 14] as (0) Non-existent, (1) Initial/Ad-Hoc, (2) Repeatable but Intuitive, (3) Defined, (4) Managed and Measurable, and (5) Optimized [8]. The maturity of processes is analyzed in the capability dimension, but not in the coverage and control dimensions. However, COBIT provides powerful controls for measuring processes both internally and externally through process and activity metrics and goal fulfillment. These concepts can be leveraged for preservation processes [3].

A recent analysis in the DP domain applied IBM's Component Business Model approach to relate DP-related business components to business areas with common objectives and evaluated the alignment of organizational structures with changing requirements of collections management and digital preservation [30]. The first SHAMAN Reference Architecture (SHAMAN-RA) presented in [2] has strong foundations in EA. However, it does not explicitly take existing domain knowledge and reference models into account in a degree sufficient to enable their transparent convergence. Based on these observations, recent work accommodated and explicitly expressed DP domain knowledge in the framework of an established Enterprise Architecture approach [1] and integrated DP capabilities with IT Governance [3]. The work presented here advances this by introducing a detailed capability model for preservation capabilities, specifying ca-

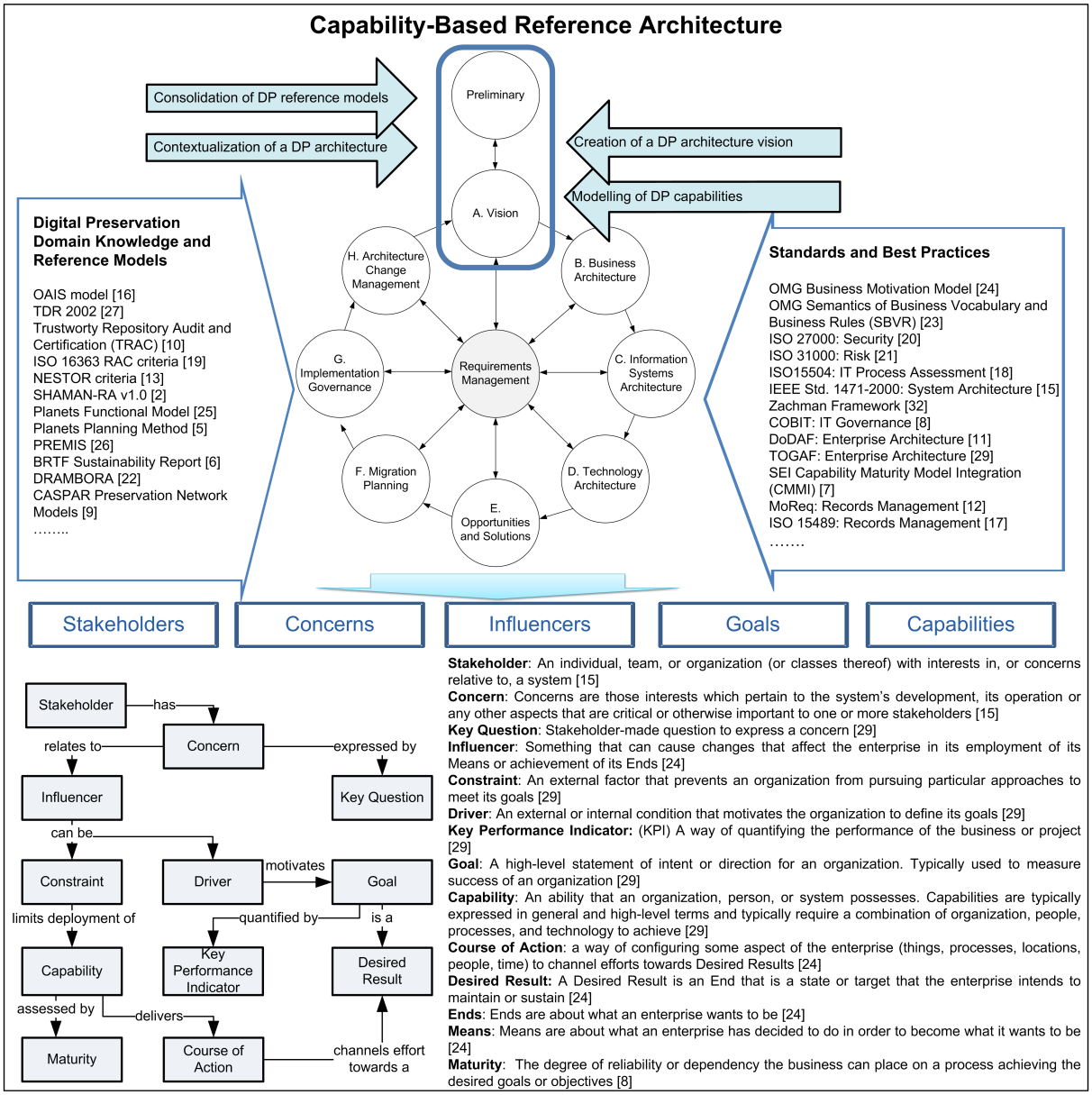

Figure 1: Using TOGAF to blend Reference Models into the SHAMAN Reference Architecture.

pability maturities for operative preservation, and analyzing the relationships between compliance criteria, drivers, stakeholders' concerns, and capabilities.

3. DIGITAL PRESERVATION CAPABILITIES

The main goal of a *Reference Architecture* is to provide a process from which concrete architecture artifacts can be derived [2]. The architecture described in [1, 3] is strongly based on TOGAF and combines it with key concepts of the Business Motivation Model (BMM) [24]. It is centered on the concept of capabilities. Note that a *capability* is fundamentally different from a system function or a process. It is instead viewed as a goal-oriented concept. A capability in TOGAF is an 'ability that an organization, person, or system possesses. Capabilities are typically expressed in general and high-level terms and typically require a combination of organization, people, processes, and technology to achieve' [29].

A successful architecture has to correctly reflect the concerns of the stakeholders of the system, from end users to developers, providing answers to whatever pertinent questions they might have. Typical digital preservation concerns include diverse aspects such as system end-usage, management, compliance, operations, and solutions.

A capability model for DP needs to be inherently independent of the business domain and, in particular, independent of the organizational scenario in which DP is deployed. It should be applicable equally to a traditional archival organization as to a business organization that is adding DP as a support capability to its primary business capabilities. It should further support organizations in answering critical questions such as *'What is the impact of a certain regulatory constraint? How can it be addressed?'* and *'How can*

we assess our processes and abilities against best practices? How can we develop targeted strategies for improvement?'.

The TOGAF Architecture Development Method (ADM) is the core component of TOGAF. It provides a systematic framework for developing an enterprise architecture. It is centered around requirements management and provides a continuous process for addressing contextual concerns and changing requirements to ensure the organization's business and IT needs are met.

We leverage the ADM to accommodate domain-specific concerns represented in DP knowledge bases and reference models. Following the ADM's first two phases, *Preliminary* and *Architecture Vision*, this requires a number of analytical steps to consolidate DP reference models, contextualize a DP architecture, model DP capabilities, and create a DP architecture vision [1].

Figure 1 illustrates the key elements of the Reference Architecture. The cyclic ADM workflow picture in the top center serves as the catalysator process into which DP domain knowledge and reference models are fed. These provide the architecture context [1], guided by standards and best practices in areas such as Information Systems; GRC; Organizational Engineering; Enterprise Architecture; and Software Engineering. Additional sources were considered, but space constraints prevents a full discussion of domain knowledge sources and their representation on the diagram. The result of our analysis is a capability-based Reference Architecture for DP that relates stakeholders and their concerns to the relevant drivers and constraints, and connects this to desired goals and required capabilities. The core concepts and their definitions and relationships are given in the bottom of Figure 1. The Reference Architecture can be used to derive concrete architectures in diverse scenarios where DP is of concern. For any concrete instantiation, additional situation-specific concerns are integrated and reconciled to produce a specific architecture by relying on the ADM process model. While previous discussions of this model focused on the high-level relations between the capabilities and their integration within an organization [1, 3, 4], we will focus here on the detailed component capabilities of preservation and specify a maturity model for preservation operations. We further outline performance measures that can be used to assess the maturity and performance of organizational capabilities along a number of dimensions.

From the analysis of the DP references, several stakeholders were identified. Stakeholders with end-usage concerns include the *Producer/ Depositor* and the *Consumer* stakeholders, which are identical in definition to the OAIS *Producer* and *Consumer* roles. The *Producer/ Depositor* stakeholder is the entity responsible for the ingestion of the objects to be preserved. Typically, its concerns include: the deposit of objects along with whatever additional data required, in accordance with negotiated agreements/contracts; assurance of access rights to the objects; assurance of the authenticity of provenance of the deposited objects; and preservation of the objects and associated rights beyond the lifetime of the repository. The *Consumer* stakeholder represents users accessing the preserved objects, with a potential interest in its reuse and a certain background in terms of knowledge and technical environment. Its concerns include the access to the preserved objects in accordance with negotiated agreements/contracts, and the correspondence of the retrieved content to its needs in terms of understandability and authenticity. Other identified stakeholders include *Management*, a generalization of all management stakeholders concerned with ends and means. Specializations of the *Management* stakeholder include the *Executive Management*, *Repository Manager*, *Technology Manager*, and *Operational Manager*. Stakeholders with compliance concerns include the *Regulator* and the *Auditor*. Operational concerns are shared between the *Repository Operator* and *Technology Operator*. Finally, stakeholders with solutions-related concerns include the *System Architect* and the *Solution Provider*.

The analysis of stakeholders' concerns, typical compliance requirements, domain models and other sources of knowledge enables an analysis of the main influencers that have an impact on the setting of organizational goals in digital preservation. Such influencers can be either drivers or constraints. The key distinction made between these influencers is between *internal* and *external* influencers. These influencers in turn drive and constrain an organization's definition of high-level goals, i.e. the desired results that an organization wants to achieve. Such goals strongly relate to stakeholders' concerns such as the user community's perception of content's authenticity, and require certain abilities inside the organization to achieve corresponding outcomes. A detailed discussion and categorization of DP drivers, an assessment of possible constraints (through external drivers), and an analysis of exemplary DP goals and their associated Key Performance Indicators is described in [1].

The organization's stakeholders, concerns, and goals in turn drive the clarification of its value chain definition and, finally, the specification of the abilities that it needs to achieve its stated goals. Figure 2 shows the high-level capability model. Capabilities are grouped into governance capabilities, business capabilities and support capabilities. In general, governance capabilities control business and support capabilities; business and support capabilities inform governance capabilities; and business capabilities depend on support capabilities. These high-level capabilities are described in [4]. The core business capability of DP in this model is **Preserve Contents** – the 'ability to maintain content authentic and understandable to the defined user community over time and assure its provenance' [1]. This is at the heart of DP, it addresses the core requirement of authenticity, understandability and provenance. This core capability is composed of two capabilities: Preservation Planning and Preservation Operation. **Preservation Planning** is 'the ability to monitor, steer and control the preservation operation of content so that the goals of accessibility, authenticity, usability and understandability are met with minimal operational costs and maximal (expected) content value. This includes managing obsolescence threats at the logical level as the core risk affecting content's authenticity, usability and understandability'[3].

Preservation Planning consists (at a minimum) of the capabilities **Planning Operational Preservation** and **Monitoring**. *Planning Operational Preservation* is the ability to make drivers and goals operational, i.e. define objectives and constraints represented by decision criteria, and assess options against these criteria to deliver efficient decisions and operational plans. It is composed of a number of component capabilities:

1. *Influencers and Decision Making*: The ability to make drivers and goals operational, i.e. define objectives and constraints represented by decision criteria, and

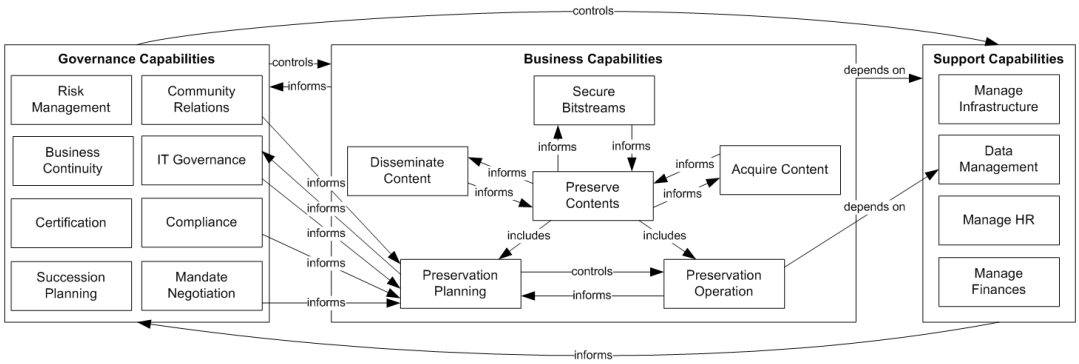

Figure 2: Capability Relations for *Preserve Contents*. Relations can be *inform, control, include, depend on.*

assess options against these criteria to deliver efficient decisions and operational plans.

2. *Options diagnosis*: The ability to gather information about available options, i.e. measures corresponding to a set of criteria.

3. *Specification and Delivery*: The ability to specify actions and directives in an understandable form and deliver it to operations (to prepare the deployment of plans).

The second planning capability, *Monitoring*, is the ability to monitor operations (in particular the execution of plans) and the environment, i.e. the ability to monitor all influencers having a potential impact on plans to ensure conformance of results to expected outcomes and notify the decision making capability (Planning Operational Preservation) of a change that requires assessment. It is composed of

1. *Internal Monitoring*: The ability to monitor operations for certain properties of interest, which include operations specified by plans and operational attributes of the system, i.e. internal influencers. Internal influencers of interest include (but are not necessarily limited to) operational statistics about the execution of delivered plans, operational statistics about content dissemination, and user feedback about satisfaction with respect to user access requirements.

2. *External Monitoring*: The ability to monitor external influencers of interest through the corresponding governance capabilities. External influencers include (but are not necessarily limited to): Technological opportunities for improving achievement of goals (through IT Governance); Technology correspondence (as an indicator of impending obsolescence, misalignment to user access requirements) (through IT Governance, Disseminate Content, Acquire Content); User community shifts (through Community Relations); Access requirements (through Disseminate Content); and Regulations (through Mandate Negotiation, Compliance).

Preservation Operation is 'the ability to control the deployment and execution of preservation plans. This includes analysing content, executing preservation actions and ensure adequate levels of provenance, handling preservation metadata, conducting Quality Assurance, and providing reports and statistics, all according to preservation plans. Preservation Actions are concrete actions (usually implemented by a software tool) performed on content in order to achieve preservation goals. For example, a migration of content to a different format using a certain tool in a certain configuration and environment'[3]. Preservation Operation is in turn composed of six component capabilities:

1. *Analysis:* The ability to measure properties of relevance in the content and document them in understandable form.

2. *Action*: The ability to execute preservation actions in order to actively preserve content according to preservation plans

3. *Quality Assurance*: The ability to deliver accurate measures that quantify the equivalence of performances (renderings) of preserved content by measuring properties of renderings/performances and comparing them to each other to measure their equivalence corresponding to requirements.

4. *Preservation Metadata*: The ability to read, understand and write appropriate preservation metadata corresponding to chosen standards.

5. *Plan Deployment*: The ability to receive plans from Planning and deploy them to an operational environment.[2]

6. *Reporting and Statistics*: The ability to produce documentation of activities in an adequate and understandable form (for monitoring and auditing).

Each of these component capabilities can be measured along a number of metrics. *Options Diagnosis*, for instance, can be measured along the following dimensions:

- Completeness: Measures are delivered for all options and each criterion.
- Correctness: All measures are correct.
- Timeliness: All measures are delivered in a certain time frame.

Similarly, the Monitoring capabilities can be tracked for completeness, correctness, timeliness and currentness.

On the operations side, performance indicators for *Actions* will include

- Completeness: Successful execution of all actions deployed as part of a plan.

[2]Technically, this may result in a set of operations potentially combining analysis, actions, QA, metadata, and reporting, all of which may be specified by the plan. Execution of the plan may require a combination of services, orchestration, and processes involving human intervention.

- Provenance: Delivery of complete audit trails to ensure provenance for every action executed.
- Results documentation: Delivery of complete information about the correspondence of action results to expected results.
- Operations documentation: Delivery of complete information about the state of operations at any point in time.

Metrics for *Reporting and Statistics* will generally include the following.

- Timeliness: Reports and statistics are delivered within a certain time frame after requested.
- Currentness: Reports and statistics always show up-to-date information, i.e. delay is below certain threshold.
- Completeness: Reports and statistics contain all relevant information about all operations.
- Relevance: Reports and statistics contain minimum unnecessary information.
- Correctness: Information reported is correct.
- Understandability: Reports and statistics are understandable by all consuming entities.

Clearly, the exact metrics that are available and meaningful in a concrete environment will depend on the organizational processes and tools available. Furthermore, the metrics described above are oriented towards an internal measurement of capabilities, and as such need to be complemented by external measures related to goal achievement. For example, the core goal of delivering authentic, understandable, and usable content to the user community can be associated with a KPI such as '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' [1]. A specification of the relationships between these process metrics and the associated outcomes of capabilities measured in KPIs is needed to achieve full control over preservation processes. However, apart from goal achievement and process metrics, capabilities can also be analyzed on a more abstract level for their maturity.

4. A MATURITY MODEL FOR PRESERVATION OPERATIONS

Focusing on strategic process and capability improvement rather than formal certification of processes, COBIT provides maturity level specifications for each process along a number of dimensions similar to [18]. We can thus assess the maturity of the *Preservation Operation* capability on the dimensions (1) *Awareness and Communication*, (2) *Policies, Plans and Procedures*, (3) *Tools and Automation*, (4) *Skills and Expertise*, (5) *Responsibility and Accountability*, and (6) *Goal Setting and Measurement*.

Table 1 defines criteria for the *Preservation Operation* capability for each maturity level and dimension. Similar criteria have been specified for Preservation Planning elsewhere [3]. As an illustrative example, consider an organization with the following diagnosis on their preservation operations: *Management is aware of the role of operations for authenticity and provenance, and there is a defined process for operations. This process includes all activities (actions, analysis, Quality Assurance, Metadata, and Reporting), and it relies on standardized plans. These plans are generally deployed according to specifications, but the deployment and operation is a mostly manual process of initiating operations as far as they are concretely specified by these plans. QA and metadata management is not driven by plans, and it does not seem to be aligned with business goals. There are guidelines about statistics and reporting procedures, but no integrated system exists for tracking the state of operations and the results of actions, and no formal metrics have been defined. Several automated tools are employed in different processes. However, the processes and rules used are defined by the availability of components and services and the level of skills of the people running these processes. A formal training plan has been developed that defines roles and skills for the different sets of operations, but all training is in fact still based on individual initiatives and not continuously managed.*

Assessing the organization's capability along the dimensions outlined above, it can be considered to be on the *Defined* level for all dimensions. Considering the skills and expertise set in the example above, we can verify that staff has operational skills and a formal training plan was developed. The absence of formal responsibility and accountability plans, however, increases the organization's dependency on specific people, which increases the severity of losing key staff trained on individual initiatives and not continuously managed. Notice that in reality, processes will generally be on different maturity levels for varying dimensions [3]. Awareness and Communication, for example, often precedes automation and tool support.

This type of capability assessment provides an internal benchmarking of the quality of processes in several dimensions. The analysis provides organizations with a decision support mechanism to prioritize actions to improve the quality of their capabilities (what and how can be improved). On the other hand, we must recognize the existence of dependencies between distinct capabilities, as shown in Figure 2. For instance, Preservation Operation *informs* Preservation Planning, but depends on other capabilities. Thus, to systematically improve the performance and maturity level of specific capabilities, we also need to consider the quality of related capabilities and understand the dependencies and the relations between internal process metrics and external outcome indicators.

5. ANALYZING CONSTRAINTS, GOALS AND CAPABILITIES

When an organization intends to analyze the impact of policies, external influencers and regulatory compliance constraints, it is often unclear which areas are concerned, and how to represent the impact (and measure the fulfillment) of certain influencers. In particular the interplay between drivers and constraints and their accumulated impact on required processes and functions is difficult to assess.

A core strength of an EA-based approach is the clear definition and separation of concerns and the traceability that it provides for impact assessment of changes. Arising constraints and drivers can be assessed with respect to the effects that they cause on concerns, goals and capabilities. Relying on the conceptual model outlined above, these can thus be addressed along the following dimensions.

	Awareness and Communication	Policies, Plans and Procedures	Tools and Automation	Skills and Expertise	Responsibility and Accountability	Goal Setting and Measurement
1	Management recognizes the need for preservation operations. There is inconsistent and sporadic communication.	Some operations are carried out, but they are not controlled. No useful documentation is produced about procedures and actions.	Some tools may be employed by individuals in an unsystematic ad-hoc manner.	There is no common awareness of which skills and expertise are required for which tasks.	There is no common awareness of responsibilities.	There is no clear awareness of goals; operations solely react to incidents and are not tracked.
2	Management is aware of the role of operations for authenticity and provenance. No formal reporting process exists, but there is some documentation about process results. Reports are delivered by individuals.	Some operational procedures emerge, but they are informal and intuitive. Operations rely on individuals; different procedures are followed within the organization. QA is recognized as a process, but mostly carried out ad-hoc and manual.	Automated tools are beginning to be employed by individuals based on arising needs and availability. Their usage is unsystematic and incoherent.	Staff obtain their operational skills through hands-on experience, repeated application of techniques and informal training by their peers.	Responsibility for operations emerges, but is not documented. Accountability is not defined.	There is individual awareness of short-term goals to achieve in operations, but no consistent goal definition or measurement.
3	Management understands the role of operations for authenticity and provenance. There are guidelines about statistics and reporting procedures, but they are not consistently enforced.	There is a defined process for all operations that relies on standardized plans. The processes and rules used are defined by available components, services and skills. QA and metadata management are not driven by business goals.	Plans are deployed according to specifications, but the process of initiating operations is mostly manual. No integrated system exists for tracking the state and results of operations.	A formal training plan has been developed that defines roles and skills for the different sets of operations, but formalized training is still based on individual initiatives.	Responsibility for operations is assigned, but accountability is not provided for all operations.	Operational goals are specified, but no formal metrics are defined. Measurements take place, but are not aligned to goals. Assessment of goal achievement is subjective and inconsistent.
4	Management fully understands the role of operations for authenticity and provenance and how they relate to business goals in the organization. Reporting processes are fully specified and adhered to.	Plans are fully deployed as operational activities, and the compliance of all operations to goals and constraints specified in plans is fully monitored. All Operations are actively monitoring state of operations.	An automated system exists to control automated operations, and automated components are widespread, yet not fully integrated.	Required skills and expertise are defined for all roles, and formal training is in place.	Responsibility and accountability for all operations is clearly defined and enforced.	A measurement system is in place and metrics are aligned with goals. Compliance monitoring is supported and compliance enforced in all operations.
5	Operations are continuously improving. An integrated communication and reporting system is fully transparent and operates in real time.	Extensive use is being made of industry good practices in plan deployment, analysis, actions, metadata, QA, and reporting.	All operations are fully integrated, status is constantly available in real-time.	Operators have the expertise, skills and means to conduct all operations. Continuous skills and expertise assessment ensures systematic improvement.	A formal responsibility and accountability plan is fully traceable to all operations.	Compliance is constantly measured automatically on all levels. Continuous assessment drives the optimization of measurement techniques.

Levels: 1: Initial/Ad-Hoc, 2: Repeatable but Intuitive, 3: Defined, 4: Managed and Measurable, 5: Optimized [8]

Table 1: Maturity Levels for the capability *Preservation Operation*

- *Stakeholders concerned*: Which are the stakeholders whose interests and viewpoints are affected by the influencer? How will it change their view of the world?
- *Concerns addressed*: Which concerns will need to consider the exact implications of the influencer? Do the Key Questions accurately reflect these considerations? Is it possible to model the influencer and its impact in the defined viewpoints and perspectives that represent the concerns?
- *Drivers involved*: Which organizational drivers are involved? What is the combined effect of a regulatory constraint and a business driver on the organizational goals?
- *Goals impacted*: Which organizational goals may be affected by an influencer, and how?
- *Capabilities affected*: Which capabilities will need to consider the effect of the influencer in order to be successfully achieving their stated goals? How can they accommodate this influencer?
- *Metrics applicable*: Which Key Performance Indicators need to be tracked to detect the exact effect of an influencer on the organization's achievement of goals? Which metrics can be used to assess capabilities? How mature are our capabilities?

Consider the case of RAC 4.1.1, *The repository shall identify the Content Information and the Information Properties that the repository will preserve.* This is part of 4.1 *Ingest: Acquisition of Content.* Based on the capability-centered Reference Architecture, it becomes possible to analyze the impact of a regulatory or organizational constraint along the lines outlined above:

- *Stakeholders concerned*: The primary stakeholders concerned include Producer/ Depositor; Consumer; and Management. However, the Repository Operator and the Solution Provider may be involved, depending on the organization's process model and the decisions taken by Management.
- *Concerns addressed*: Focusing on the OAIS-related stakeholders, the concerns addressed include (Key Questions in brackets):
  1. *Producer/Depositor: Authenticity and Provenance.* Content provided is authentic and has complete provenance. (What kinds of guarantees will the repository provide to assure me the authenticity and understandability of my objects? Will complete provenance information be provided with the disseminated content, so that the provided objects be traceable to the original?)
  2. *Consumer: Content.* The information retrieved is authentic, understandable and corresponds to my needs. (Will the domain knowledge that I have be sufficient to access and understand the content? Will the objects be corresponding to my queries, authentic, compatible to my technical environment, and understandable?)
  3. *Management: Mandate, Mission, Policies and Compliance.* The governance of the mandate, the commitment of the organization to digital preservation, may it be for business needs, legal, or legislative reasons; and corresponding compliance. This includes certification and succession planning. (Is the mandate adequate, well-specified and appropriately accessible? Is the organization able to fulfill the mandate? Does the organization possess all the required contracts regarding succession planning and escrow agreements? Is the organization compliant to external regulations?

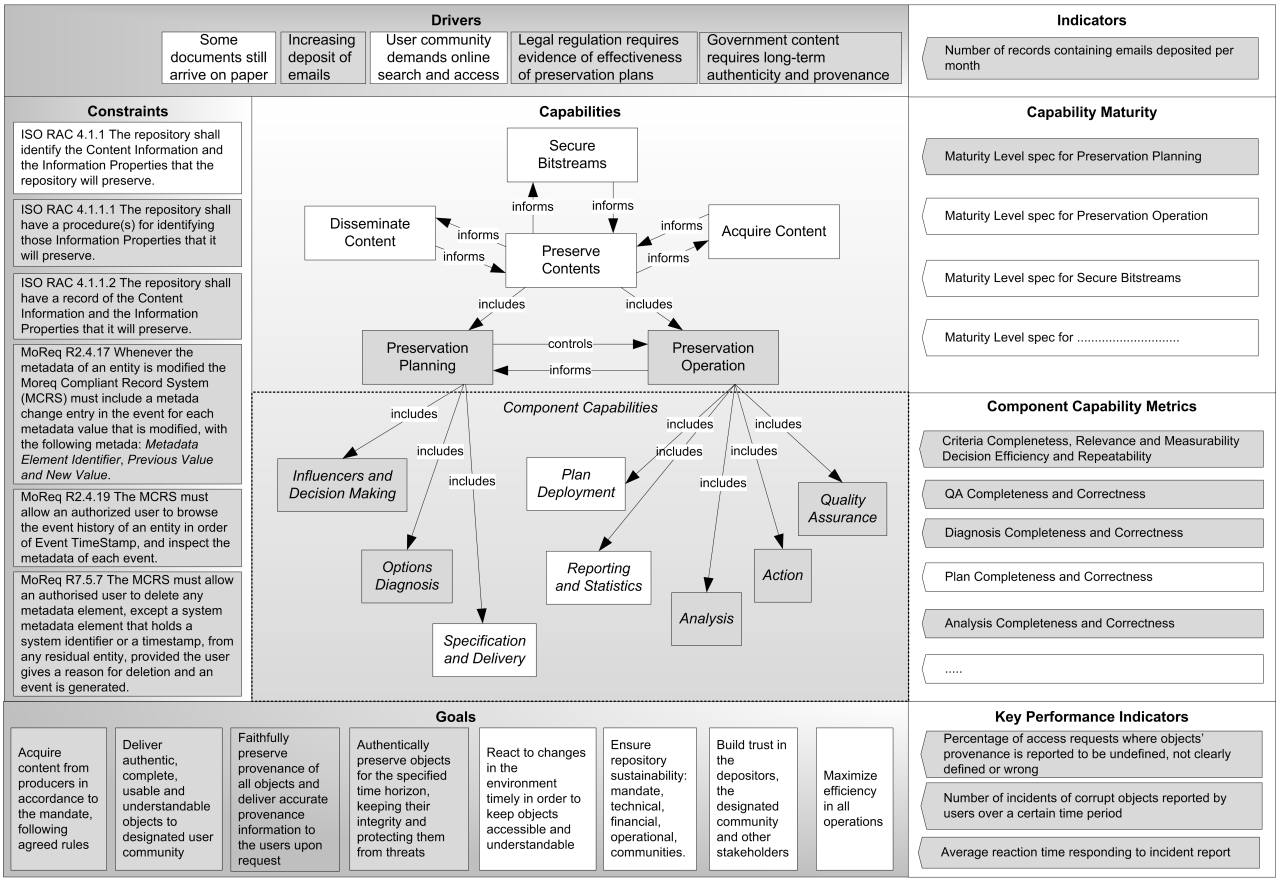

Figure 3: Business drivers and compliance to ISO RAC and MoReq2010: Content Acquisition at the CML.

Does the organization possess necessary certifications?)

- *Drivers involved*: The external drivers and constraints involved include (at least) access contracts; deposit contracts; supplier contracts; and the user community's knowledge, technology, and demand satisfaction. The internal drivers include organizational capabilities, resources (staff expertise and qualifications, existing software and costs) and the business vision.

- *Goals impacted*: The combined impact of these on the goals depends on the business vision, but will have at least an impact on the targeted level of fulfillment in terms of authenticity.

- *Capabilities affected*: Correspondingly, a number of capabilities will be affected: For example, the capability *Community Relations* may be required to consult and negotiate with communities about levels of information properties preserved. Similarly, *Preservation Planning* will need an understanding of the information properties to preserve, a reliable method for assessing the fulfillment of the goals derived from these, a method for evaluating potential ways of preserving all properties, and the ability to specify them for operational purposes. *Preservation Operation*, in turn, will need appropriate means for Analysis, Actions, and Quality Assurance that are aligned with the content that the archive has to deal with.

- *Metrics applicable*: Finally, the metrics that can be tracked can be deduced from the component capabilities affected. For example, they will include the metrics mentioned in Section 3, such as completeness and correctness of Options Diagnosis. Furthermore, external KPIs can be used to measure outcomes, and capabilities can be assessed for their maturity levels. All of these metrics can be used to assess compliance to the original regulatory constraint as specified in RAC, and used as targets to improve organizational capabilities.

As a simplified visual illustration, Figure 3 shows a real-world case where RAC 4.1.1 intersects with a business driver. The municipality of Lisbon (CML) is in the process of integrating the software Documentum[3] with a set of business workflows for a wide range of organizational entities, including the Municipal Archives. In this process, Records Management concerns overlap with DP concerns and a number of specific drivers of organizational change. Generic high-level DP goals as outlined in [1] are pictured at the bottom. Relevant business drivers to be addressed are shown on the top. These include long-term authenticity and provenance, as well as a need for evidence-based proof of effectiveness. Selected constraints posed by RAC and Moreq2010 are listed on the left. The right side shows indicators that can be tracked externally and internally to exercise control based on a quantative assessment. The related constraints, drivers, capabilities, maturities, capability metrics, goals, and KPIs are shaded in gray.

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

	Capabilities	4.1 Ingest: Acquisition of Content	4.2 Ingest: Creation of the AIP	4.3 Preservation Planning	4.4 AIP Preservation	4.5 Information Management	4.6 Access Management
Governance	Compliance	A	A	A	A	A	A
	Community Relations	S	S	S		S	
	Certification						
	Mandate Negotiation			A			
	Business Continuity						
	Succession Planning						
	IT Governance	A	A	A	A	A	A
	Manage Risks	A	A	A	A	A	A
Business	Acquire Content	R	R				
	Secure Bitstreams	S	S		S		
	Preserve Content						
	- Preservation Planning	S	S	R	S		
	- Preservation Operation	S	S		R		S
	Disseminate Content					S	R
Support	Data Management		S		S	S	
	Manage Infrastructure						
	Manage HR						
	Manage Finances						

Table 2: High-level capabilities (A)ware of, (R)esponsible for or (S)upporting RAC criteria in group 4

The increasing move towards email deposit intersects with RAC 4.1.1, since the significant properties that will be preserved need to be decided. This is a typical task for preservation planning, which will require a clear documentation of decision factors and the ability to diagnose possible options for email preservation to decide on a feasibility and level of authenticity that can be guaranteed. On an operational level, this requires processes and tools for email analysis and quality assurance for potential preservation actions. The affected component capabilities can be assessed along the measures outlined above, while Preservation Planning and Preservation Operation can be assessed for capability maturity. On the level of end-user results, i.e. business outcomes, Key Performance Indicators can be used to track goal achievement from an external perspective.

Table 2 summarizes the impact of each group of RAC criteria in section 4 (Digital Object Management) on the capabilities. Essentially, a criterion can be (part of) the primary *responsibility* of a capability, or a capability may be indirectly required to support the fulfilment. For example, operational verification of content integrity as requested in RAC 4.2 – which is primarily concerned with Ingest – requires fixity checks, which are part of Data Management. Finally, certain capabilities may need to be *aware* of compliance criteria to be successful in *their* mission. For example, *Compliance* is affected by all constraints – since its mission is to 'verify the compliance of operations and report deviations' [4], it will need to be aware of all compliance constraints. This applies a priori to Governance, Risk and Compliance, but is also required in other areas. In this sense, it is interesting to see how certain groups of criteria have an impact beyond the obvious one that refers to the directly responsible capability. For example, the criteria listed in section 4.1 influence not only the *Acquire Content* capability, but also others, such as the business capability *Secure Bitstream*. This is caused by *'4.1.6 The repository shall obtain sufficient control over the Digital Objects to preserve them.'*[19], which makes direct references to bitstream preservation.

The compliance with RAC criteria will also have an impact on the maturity level of capabilities. However, this is dependent on the way that compliance is achieved. For instance, RAC 4.1.5, *The repository shall have an ingest process which verifies each SIP for completeness and correctness*, may influence the maturity levels for the *Policies, Plans and Procedures* dimension of the Acquire Content capability. Depending on the way compliance is monitored, it can also impact the *Tools and Automation* dimension, if the verification is automated. Other dimensions will be impacted as well, although indirectly.

6. DISCUSSION AND OUTLOOK

The Reference Architecture that forms the basis of this article is an Enterprise Architecture-based approach that enables the accommodation of digital preservation concerns in the overall architecture of an organization. For that, a capability-based model of preservation was derived from established digital preservation key references and best practices from related fields. This included in-depth analysis of the stakeholders of the domain, their concerns, goals, and influencers (drivers and constraints). The result is a multidimensional view on the domain concepts covered in these key references. The approach taken with this Reference Architecture enables the transfer of DP know-how into a nontraditional repository-based DP scenario, since it is itself agnostic to concrete scenarios. In other words, this capability-based approach can deliver value to organizations in which the preservation of contents is not a main business requirement, but required to enable actual delivery of value in the primary business.

The specification of internal process metrics and external metrics measuring the achievement of certain goals by each capability through KPIs represents an essential step towards a quantified control mechanism that can be used effectively to exercise control and govern capabilities [3].

The approach provides a powerful tool to enable responsible stakeholders to analyze the impact of compliance regulations and constraints on their systems' architecture requirements and their organizational capabilities. It can furthermore be used to assess capability maturity and process maturity to enable focused improvement of key areas. It thus enables organizations to improve maturities by considering the impact that compliance requirements have on organizations' capabilities and processes. Based on a maturity assessment, an organization can target a *capability increment* to improve its capabilities and their maturities by undergoing a change initiative to increase performance for a particular capability [29].

Current work is focused on moving forward in the TOGAF-

ADM cycle to derive a contextualized *Business Architecture* for a concrete real-world scenario, and conducting a full-depth analysis of the combined implications of constraints coming from the domains of DP and Records Management in a real-world case. This furthermore sets the grounds for a full maturity model on all capabilities.

Acknowledgments

This work was supported by FCT (INESC-ID multi-annual funding) through the PIDDAC Program funds and by the projects SHAMAN and SCAPE, funded under FP7 of the EU under contract 216736 and 270137, respectively.

7. REFERENCES

[1] G. Antunes, J. Barateiro, C. Becker, R. Vieira, and J. Borbinha. Modeling contextual concerns in Enterprise Architecture. In *Fifteenth IEEE International EDOC Conference*, Helsinki, Finland, August 29 - September 2 2011.

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

[3] C. Becker, G. Antunes, J. Barateiro, R. Vieira, and J. Borbinha. Control Objectives for DP: Digital Preservation as an Integrated Part of IT Governance. In *Proc. 74th Annual Meeting of ASIST*, New Orleans, October 2011.

[4] C. Becker, G. Antunes, J. Barateiro, R. Vieira, and J. Borbinha. Modeling digital preservation capabilities in enterprise architecture. In *In 12th Annual International Conference on Digital Government Research (dg.o 2011), June 12-15, College Park, MD, USA.*, 2011.

[5] 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.

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

[7] Software Engineering Institute. Capability Maturity Model Integration for Development. Version 1.3. Carnegie Mellon University, November 2010.

[8] IT Governance Institute. CobiT 4.1. framework – control objectives – management guidelinces – maturity models, 2007.

[9] E. Conway, M. Dunckley, B. Mcilwrath, and D. Giaretta. Preservation network models: Creating stable networks of information to ensure the long term use of scientific data. In *Ensuring Long-Term Preservation and Adding Value to Scientific and Technical Data*, Villafranca del Castillo, Madrid, Spain, 2009.

[10] 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.

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

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

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

[14] D. L. Gibson, D. R. Goldenson, and K. Kost. *Performance Results of CMMI-Based Process Improvement*. Software Engineering Institute, Pittsburg, PA, 2006.

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

[16] ISO 14721:2003. Open archival information system – Reference model, 2003.

[17] ISO 15489-1:2001. Information and documentation: Records management, 2001.

[18] ISO/IEC 15504-1:2004. Information technology - Process assessment – Part 1: Concepts and Vocabulary, 2004.

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

[20] ISO/IEC 27000:2009. Information technology - Security techniques - Information security management systems - Overview and Vocabulary, 2009.

[21] ISO 31000:2009. Risk management – Principles and guidelines, 2009.

[22] A. McHugh, R. Ruusalepp, S. Ross, and H. Hofman. The digital repository audit method based on risk assessment (DRAMBORA). In *Digital Curation Center and Digital Presevation Europe*, 2007.

[23] Object Management Group. *Semantics of Business Vocabulary and Business Rules (SBVR), Version 1.0*. OMG, 2008.

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

[25] PLANETS Consortium. Report on the planets functional model. Pp7/d3-4, 2009.

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

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

[28] C. Rosenthal, A. Blekinge-Rasmussen, J. Hutar, A. McHugh, S. Strodl, E. Witham, and S. Ross. *Repository Planning Checklist and Guidance*. HATII at the University of Glasgow, 2008.

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

[30] R. J. van Diessen, B. Sierman, and C. A. Lee. Component business model for digital repositories: A framework for analysis. In *Proc. iPRES 2008*, 2008.

[31] 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.

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