Component Business Model for Digital Repositories: A Framework for Analysis

Raymond J. van Diessen
IBM Global Services
Johan Huizingalaan 765
1066VH, Amsterdam
The Netherlands
Raymond_vanDiessen@nl.ibm.com

Barbara Sierman
National Library of the Netherlands
Prins Willem-Alexanderhof 5
2509 LK The Hague
The Netherlands
Barbara.Sierman@kb.nl

Christopher A. Lee
School of Information and Library Science
University of North Carolina
CB#3360, 100 Manning Hall
Chapel Hill, NC 27599-3360
callee@ils.unc.edu

Abstract

Digital preservation is too big a challenge for any institution or solution supplier to confront on its own. The success of any long-term digital repository will depend upon multiple "open" services provided by a wide range of service providers. No company or organisation in the world is able to provide the preservation solution for all known formats, object types, or policies. Viable approaches are likely to span organizational, institutional and national boundaries. In 2003 the KB, National Library of the Netherlands, in cooperation with IBM, developed the e-Depot as their solution for long-term preservation of digital publications. The core of the e-Depot is IBM's Digital Information Archiving System (DIAS). This article will discuss the exercise of the KB/IBM Research Group to apply IBM's Component Business Modelling (CBM) in a digital preservation environment. The CBM map is used by a process called Goal Service Modelling (GSM) to identify candidate services for future versions of the e-Depot. Heat maps are used for impact analysis – to discuss organisational structures, existing hardware and software solutions and business processes in the context of the CBM map. The approach is suggested as a way for other repositories to manage and coordinate their activities, as well complimenting current repository audit and certification activities.

Introduction

There has been a growing professional understanding of what a trustworthy digital preservation repository should look like. The Trustworthy Repositories Audit and Certification: Criteria and Checklist (TRAC 2007), Digital Repository Audit Method Based on Risk Assessment (DRAMBORA) (McHugh et al. 2007), and Network of Expertise in Long-Term Storage of Digital Resources (NESTOR) (Dobratz et al. 2006) all identify measures that institutions should take to support a trustworthy repository, not only with ingest, but also with the other functions, such as preservation planning and access. Tools and services are now available to support many of the core functions, e.g. the JSTOR/Harvard Object Validation Environment (JHOVE) and Digital Record Object IDentification (DROID) for file characterization; kopal Library for Retrieval and Ingest (koLibRI) and Producer - Archive Workflow Network (PAWN) for ingest; Automated Obsolescence Notification System (AONS) for preservation planning; Typed Object Model (TOM) for migration; and Dioscuri and the Universal Virtual Computer (UVC) for emulation. Repositories with crucial file format information are being developed, including PRONOM, the Global Digital Format Registry (GDFR), and IBM Preservation Manager. Several international projects - including Preservation and Long-Term Access through Networked Services (PLANETS) and Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval (CASPAR) - aim explicitly to deliver tools and systems to support parts of the digital preservation process. We can expect more tools and services to be developed specifically for long-term digital preservation in the years ahead. While these tools and services provide a valuable contribution to managing digital preservation, it is up to the organisations themselves to integrate and manage these tools and services in their digital preservation environments.

The KB, National Library of the Netherlands, manages the e-Depot. At this moment, 11 million digital objects are stored in the e-Depot. Many new types of material are coming, and they will be more complex (such as websites, e-books and compound objects). In 2006 the KB and IBM formed a Research Group to develop a vision of how to integrate external tools and services into the architecture of the e-Depot. In this exercise IBM's Component Business Modelling (CBM) method played an important role. This method allows an organisation to map its strategies to relevant business components that support the organisation's objectives and helps to identify the most important business components. Each of these business components – in the case of the KB, most being departments within the library – will need services to reach

their goals. The CBM method provides a framework for viewing and analyzing the organisation as a network of individual components. Once processes and organization are dissected into discrete understandable and manageable components, the unique activities and associated resources, tools and services for each individual business component can be identified. Through the definition of the business components, the responsibilities for the management of the associated resources is clearly specified.

At the KB, the focus has been on business components related to the long-term digital preservation activities of the e-Depot. Both the actual situation and the future plans were input for the Research Group to identify necessary activities and the supporting services to accomplish the digital preservation goals. The Digital Repository CBM Map helps to determine when and where resources should be focused and how external services and solutions can be integrated.

Component Business Model

IBM has developed the CBM approach to help their clients to map business strategy to business components. Business components are the core building blocks of the organisation. A business component identifies a cluster of activities that together implement some set of capabilities which are offered through services. We will differentiate between business services, which can be provided either with or without the support of Information Technology (IT), and IT services, which are provided completely through software. When we use the term "services," it will imply both business and IT services. Business components can be managed independently, and their business and IT services can be reused across the organisation. CBM allows an organisation to identify its core business components, and understand where there are opportunities to outsource and/or cooperate with 3rd parties. An individual business component contains the activities and associated resources – such as organisational structure, people, skills and technology – to implement specific capabilities (services) needed by the organisation to achieve its goals.

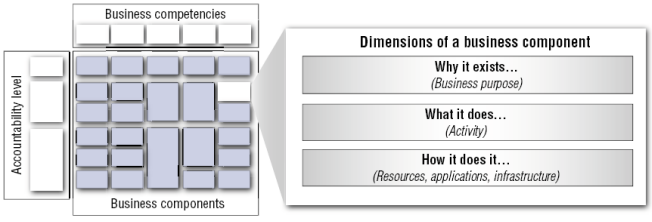

Figure 1: Basic CBM Structure

The description of every business component should involve answering three questions:

- Why it exists – What is the business purpose of the component?
- What it does – What are the core activities to support the business purpose?
- How it does it - How are the activities to be preformed and what resources are needed, e.g. people and IT support?

The business components are clustered along two dimensions. Horizontally, an accountability level characterizes the scope and intent of activity and decision-making. The three accountability levels used in CBM are Directing, Controlling and Executing:

- **Directing** - strategy, overall direction and policy
- **Controlling** - monitoring, managing exceptions and tactical decision making
- **Executing** - doing the work

Vertically, the major business competencies are identified. Business competencies are large business areas with common global objectives. For example, in a library environment, collection management and customer service management are major business competencies.

KB Objectives

In order to understand the rationale behind the identified business components, we first have to identify the strategic objectives of the KB. The mission statement of the KB identifies four major objectives:

1) We give researchers and students access to research information;
2) We enable everyone to share in the riches of our cultural heritage;
3) We foster the national infrastructure for scientific information;
4) We further permanent access to digital information within an international context.

The Research Group focused on the aspects related to Digital Preservation and especially the objective "We further permanent access to digital information within an international context."

By identifying the key business components needed to support the above objectives of the KB we create a framework for viewing the organisation as a network of individual business components. Once processes and organisation are dissected into discrete understandable and manageable components, the unique activities and associated services can be identified, along with the resources needed to execute them.

The CBM map presented in the next section has been developed with the above organisational objectives in mind. Although the focus of this article is management of digital collections, management of "paper-based"

collections will also be supported by the same set of business components.

Digital Repository CBM Map

Based on the KB´s strategic objectives mission statement discussed above, we identified five major competencies by which to cluster the individual business components:

- **Service Management**: Delivery of collection objects and associated services to the customers of the KB across the supported channels.
- **Collections Management**: Acquisition, processing and cataloguing of all publications, both for the research collection and the deposit collection.
- **Preservation Management**: Facilitating access to the different collections over-time, including addressing media decay and obsolete technology associated with each digital collection.
- **Business Management**: General management of the business of the KB.
- **IT Management**: Management of the overall IT infrastructure.

competencies. A more detailed preservation of a CBM focussed on IT management can be found in (Ernest and Nisavic 2007).

This article will not elaborate on the more generic components of the Business Management and IT Management competencies. However, we would like to stress that IT Management has some specific objects in relation to the long-term requirements of a digital repository. Not only do the digital collection assets have to be preserved for the long term, but the digital repository solution itself also should be able to adapt to technology changes. This requires the different components of such a digital repository solution to be modular with well-defined interfaces and based on open standards, as well as characterising the preservation environment itself, as it changes over time (Moore 2008).

Figure 2 presents the CBM map created for the KB, with the focus being on the management of digital collections (digital repository). Service Management addresses the services the KB uses to support its customer base. Service Management can be organized along the three CBM

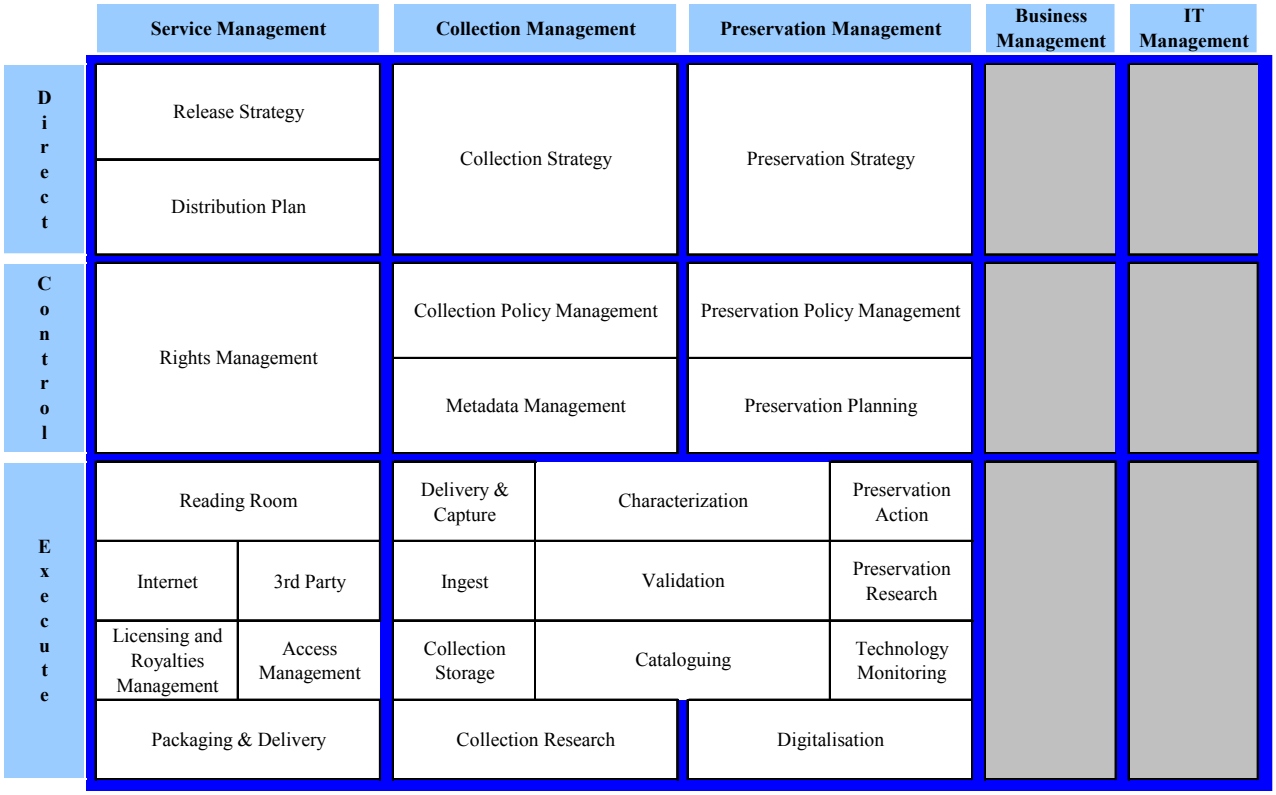

Figure 2: Digital Repository CBM Map

Service Management, Collection Management and Preservation Management are specific to organisations that manage digital collections, i.e. manage digital repositories. Business Management and IT Management are more generic and needed in any type of organisation. In this article, we have focused on the first three business

accountability levels: directing, controlling and executing. Release Strategy and Distribution Plan are part of the general strategy (Direct). Rights Management plays an essential role in determining whether and how content can be delivered to specific customer groups (Control). At the execution level are the different channels through which services are delivered.

The different Service Management business components are described below in more detail:

- **Release Strategy:** Defines which collections are available to which customer groups, and the collection enrichment services, such as abstracts and classifications, to be associated with each collection or customer group.
- **Distribution Plan:** Defines the specific access strategies to be supported for particular combinations of collections and customer groups.
- **Rights Management:** Controls potential usage restrictions to be enforced on specific collections.
- **Reading Room:** Provides services to the customers in the KB reading rooms.
- **Internet:** Provides remote services to customers via the Internet.
- **3rd Party:** Provides services related to 3rd-party organisations, such as publishers or other cultural heritage institutions.
- **Licensing and Royalty Management:** Manages all licensing, royalty and accounting aspects associated with a particular collection.
- **Access Management**: Provides mechanisms to enforce the particular licenses and rights associated with any given collection, as well as the identification and authorization of individual users.
- **Packaging and Delivery**: Prepares the selected collection objects for delivery over a selected channel to a particular customer.

Collection Management defines the business components needed to define and manage the collections of the KB. Individual collections are managed according to defined collection policies. Different categories of metadata are used to manage the collections and support access, e.g. bibliographic, archival and technical metadata. The different Collection Management business components are described below in more detail:

- **Collection Strategy**: Decides which collections to build up and defines the value of the different collections for the KB.
- **Collection Policy Management**: Defines the rules and guidelines for submissions of assets into a particular collection.
- **Metadata Management**: Identifies the different categories of metadata to be associated with particular collections and builds ontologies over the different metadata specification approaches being applied.
- **Delivery and Capture**: Pre-processes digital assets to be ingested: receives or captures digital assets and stores them in a working space for verification in conformance with the defined collection policies.
- **Ingest**: Checks the collection asset for compliance and completeness followed by the archiving of the asset.
- **Collection Storage**: Stores collection assets within the library. In the case of digital material, an Archival Information Package (AIP) has to be maintained in one or more storage environments as identified in the Reference Model for an Open Archival Information System (OAIS) (ISO 14721:2003).
- **Collection Research**: Extends knowledge and best practices for the development of collections and associated services, including access and presentation.

Preservation Management includes all the business components involved in long-term preservation of digital collections. This is still an active area of research internationally. Two sets of activities are of major importance within this competency. First, one needs to monitor the impact of technology changes over time on the different collections managed, as part of the Preservation Watch function. Second, preservation actions have to be defined to counteract the impact of technology obsolescence, either by migrating collection assets to new formats, emulating obsolete technology or a combination of both. The Preservation Management business components are described below in more detail:

- **Preservation Strategy**: Defines the preservation strategies to be supported by the KB digital repository environment.
- **Preservation Policy Management**: Specifies the preservation policies associated with particular collections or types of digital assets.
- **Preservation Planning**: Defines the actions to implement a specific preservation policy.
- **Preservation Action**: Carries out activities needed to preserve particular collections or types of digital assets: migration (converting collection assets into new formats), emulation (providing new environments to emulate obsolete environments) or a combination of both. Normalisation, i.e. transforming digital assets into formats optimised for the management of particular collections, is also a preservation action.
- **Preservation Research**: Conducts research in the field of digital technologies, network information and the preservation of digital heritage.
- **Technology Monitoring**: Monitors changes in technology environments to be addressed by preservation planning and preservation action.
- **Digitisation**: Specialised preservation action that converts analogue assets into digital assets.

Three business components have been identified as being important to both Collection Management and Preservation Management. These business components are related to the characterisation, validation and cataloguing of collection assets. They are used for initial ingest of a collection but also provide activities which are important for the preservation of the collection. The business components are described below in more detail:

- **Characterisation**: Identifies and records the important characteristics of a digital object and facilitates

searching across the characterisation metadata.

- **Validation**: Checks whether the collection assets conform to the associated collection and preservation policies e.g. conformance to file format specifications.
- **Cataloguing**: Builds and maintains metadata to facilitate both general and domain-specific access and searching within or across collections.

Applying the Digital Repository CBM Map

There are a number of ways that the Digital Repository CBM map can be applied within an organisation. The development of the CMB map was triggered by the objectives of KB/IBM Research Group to look into the requirements for an open and integrated preservation framework for the KB to extend the e-Depot, based on IBM´s DIAS solution. The Research Group was aware that any durable electronic deposit solution can never be dependent upon a single vendor providing a closed solution. New formats and preservation tools will continue to be introduced over time, requiring any given solution to be sufficiently open to incorporate functionality from 3rd parties.

The next section will show how the Digital Repository CBM map has been used to identify the generic services to support future developments of the e-Depot. We will then explain how the Digital Repository CBM map can be used to facilitate impact analysis with regard to IT and organisational support.

Identifying Services

The CMB map provides the top-down starting point for the identification of the services that need to be provided by individual business components. Each business component has its own business activities, which can be performed manually or with the support of IT services. The business components drive the definition of potential service candidates. These service candidates are tested for functional usability by determining how they can be used in the different business processes. In order to ensure that a candidate service is reusable across various contexts, it is important to validate it against many different business processes.

This approach will also identify potential "white spots," i.e. business components not yet supported by any services. White spots are not always a problem. Some business components (e.g. Collection Strategy) might not be implemented through IT services but are, instead, based on human processes resulting in vision documents and associated implementation plans. The strength of the Digital Repository CBM map is the ability to condense major aspects of the library environment into a simple overview that is easy to communicate.

The above top-down approach does not take into account potential existing IT solutions that could provide some of the required services. The KB has already invested a large amount of money and effort in their current e-Depot solution. Therefore, we also need to evaluate how current IT solutions can provide some of the required services, i.e. bottom-up approach.

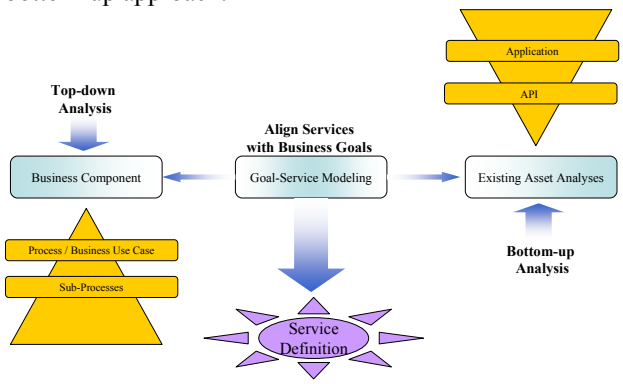

Figure 3: Goal Service Modelling (GSM)

"Goals Service Modelling" (GSM), combines the top-down and bottom-up approaches. Services are identified by the CBM Maps as well as by looking at the functionality of the existing IT solutions. The end result of this exercise is the definition of a complete set of services needed by the organisation to reach the digital preservation goals of its business components, divided in a set of existing services and needed services

Example. One example is the process called *Prepare Content Package for Ingest*. For our exercise we designed this process as in the figure 4. A similar process currently exists in a basic form and creates the Submission Information Package (SIP) to be ingested by DIAS.

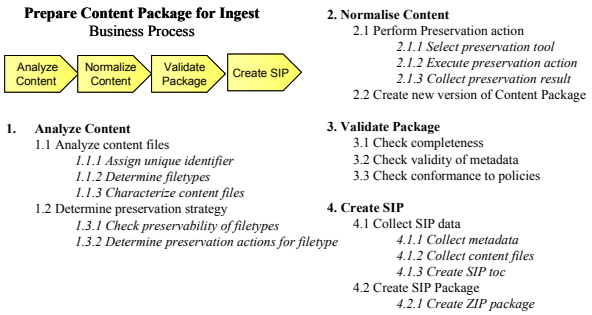

Figure 4: Services - Prepare Content Package for Ingest

Prepare Content Package for Ingest process assumes that Content Packages have been extracted from the Producer Submission Package (PSP). This package is mostly optimized to submit batches, rather than individual digital objects. For example, scientific articles are delivered in batches containing multiple articles of an issue of a periodical, but each article is then ingested individually. As a first step, the content of the packages is analyzed.

Each content file, as found in the SIP, will be given a unique identifier to trace it through the following processes. For each file, an initial determination of file type is generated, and based on this information and the associated policies for the collection a decision is made whether this file needs further characterisation. A conceptual overview of the individual activities is given in figure 4.

After analysing the content file and based on the file type information and possibly other details about the file, as gathered during characterisation, a preservation strategy is determined. First of all, the "preservability" of the file types is determined (whether they are in known and accepted formats, and known risks associated with attempting to preserve given formats), and based on this preservability check, it is determined what preservation actions are needed. An example of such action could be to normalise the file into a preferred file format, such as PDF/A (ISO 19005-1:2005).

The services needed in the business processes will all be attributed to particular business components in the Digital Repository CBM map. For instance, *Determine Filetypes* is a service of the business component Characterisation, and *Collect Metadata* is a service of the Metadata Management business component.

As discussed earlier, there are initiatives across the globe to develop services, which will benefit digital preservation. But how does one determine when a service is viable for an organisation? Which measures are needed to implement services in a manageable way? IBM has developed a Litmus Test to identify viable services, based on the experience of service-oriented architecture (SOA) implementations (where the main goal is to implement services in a flexible and manageable way). The Litmus Test is a set of questions, which need to be answered before implementing a service, from various points of view:

Business Alignment

- Is the organisation willing to fund the service through its lifecycle: provisioning, management, governance, and maintenance?
- Is the organisation willing to share the service internally or externally with clients or business partners?

Composability

- Is the service self-contained (can it be executed within the business components without any resources external to the business component except potential services to be supplied by other business components)?
- Is the service stateless (core operations can be executed as independent transactions)?

Externalized Service Description

- Is the service defined in a way that makes it clear what input and output are expected, and what the effects of the service will be?

Redundancy Elimination

- Can this service be used by the business stakeholders within all processes where its function is required?

After carrying out the exercises described above – creating a CBM map, identify the services via GSM and performing the Service Litmus Test – an organisation should have an overview of all the services that are needed and where they will be used, as well as which services are generic and reusable. For example in figure 4, the lower level services associated with 1.2 (*Determine Preservation Strategy*), could also be used in other preservation processes, e.g. when reanalyzing already ingested assets that need to be migrated.

Heat Maps

The Digital Repository CBM map can be used not only to identify the services, but also to discuss aspects of organisational and Information Systems (IS) architecture. Recall the definition of a business component as a clustering of business activities with common objectives, which potentially can be managed independently within the organisation.

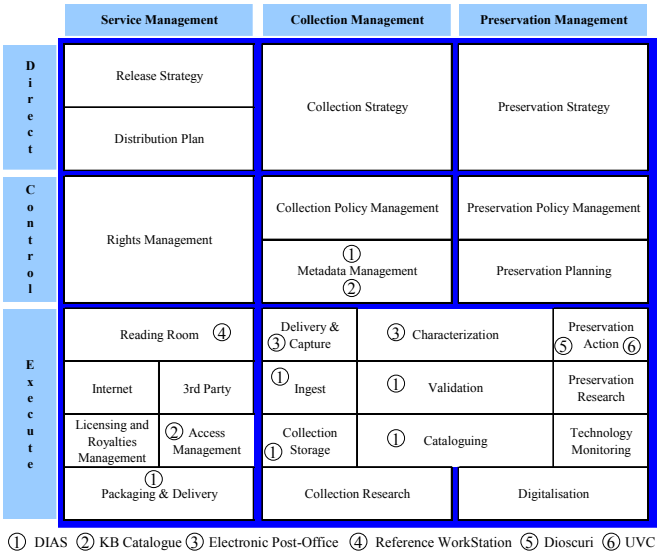

Figure 5: Example of IS Heat Map

Heat Maps are a visualization tool to map different types of needed resources onto the identified business components. They illustrate points of potential resource conflict between different business components.

The above example of a Heat Map shows how the major applications of the KB are positioned to support the business components. The IS Heat Map provides an overview of the different applications currently supporting KB´s business objectives. It also highlights potential white

spots. For example, technology monitoring at the moment seems not to be supported by any applications in the operational environment. An operational solution could be implemented in one of the next versions of DIAS based on Preservation Manager Proof of Concepts (Oltmans, Diessen and Wijngaarden 2004), using another file format registry or the implementation of the results in this area of the PLANETS project. Normally the business components would be coloured to represent their state. In the above IS Heat Map the colour would be used to represent the fit between actual and required IT support for the business component, e.g. bad, average and good.

The Heat Map approach can also be used to evaluate whether the current organisational structures are aligned with changing requirements introduced by the management of digital collections inside an organisation. Ideally, responsibility for a given business component will not be divided over multiple organisational units. The responsibility for any business component should, based on the definition, be the responsibility of one organisational unit to maintain clear responsibilities for the resources and actions to be executed by the business component. In practice, such cross-unit sharing of a component often does occur, which can generate additional risks and coordination costs.

The Digital Repository CBM map and Heat Maps can be used together to compare different organisations and digital repositories. As discussed above, several current initiatives are investigating the characteristics of trustworthy repositories, along with criteria for their audit and certification. The use of CBM, Heat maps and GSM could be used to translate these criteria into digital preservation environment solutions in specific organisational and institutional contexts.

Conclusions and Next Steps

Key to the success of any digital repository focused on long-term preservation of its collections is openness and adherence to open standards. Technology innovation is only accelerating, with new digital formats and supporting application software being introduced and digital objects becoming more complex. Digital repositories will need to adapt continuously, in order to support these new formats with appropriate services for characterisation, validation, ingest, preservation and access. At the moment, the preferred industry approach to make systems flexible is to adopt a service-oriented architecture (SOA) and associated web service standards.

The Digital Repository CBM map enables systematic analysis of the impact of new developments. We believe it is important to generalize and refine the Digital Repository CBM and discuss the results within the digital preservation community. The CMB map could also provide added value to ongoing audit and certification initiatives.

We have shown how the top-down Digital Repository CBM map provides the required context by identifying the business components needed to manage a digital repository. With GSM it is possible to identify the services. The Heat Maps show the results of impact analyses that can be facilitated by the CBM map for a multitude of factors: services, IS, resources, processes and organisation. We consider this exercise a useful method to model required services in the future.

The next steps will focus on the communication and validation of the initial Digital Repository CBM map by the long-term digital preservation community. We also want to evaluate the effectiveness of the Digital Repository CBM map in the comparison and coordination of multiple repository environments.

References

Dobratz, S.; Hänger, A.; Huth, K.; Kaiser, M.; Keitel, C.; Klump, J.; Rödig, P.; Rohde-Enslin, S.; Schoger, A.; Schröder, K.; and Strathmann, S. 2006. *Catalogue of Criteria for Trusted Digital Repositories, Version 1*. Frankfurt, Germany: nestor Working Group on Trusted Repositories Certification.

Ernest, M.; and Nisavic, J.M. 2007. Adding Value to the IT Organization with the Component Business Model. *IBM Systems Journal* 46(3): 387- 403.

ISO 14721. 2003. Space Data and Information Transfer Systems – Open Archival Information System – Reference Model.

ISO 19005-1. 2005. Document Management – Electronic Document File Format for Long-Term Preservation – Part 1: Use of PDF 1.4 (PDF/A-1).

McHugh, A.; Ruusalepp, R.; Ross, S.; and Hofman, H. 2007. Digital Repository Audit Method Based on Risk Assessment (DRAMBORA). Digital Curation Centre and Digital Preservation Europe.

Moore, R. 2008. Towards a Theory of Digital Preservation. *International Journal of Digital Curation* 3(1): 63-75.

Oltmans E.; Diessen R.J. van; and Wijngaarden H. van 2004. Preservation Functionality in a Digital Archive. In Proceedings of the Joint Conference on Digital Libraries, 279-286. Tucson, Arizona: ACM.

Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist. 2007. Chicago, Ill.: Center for Research Libraries and OCLC Online Computer Library Center.