

A Survey of Organizational Assessment Frameworks in Digital Preservation

Emily Maemura, Nathan Moles, Christoph Becker
Faculty of Information
University of Toronto

{e.maemura, n.moles}@mail.utoronto.ca, christoph.becker@utoronto.ca

ABSTRACT

As the field of digital preservation continues to mature, there is an increasing need to systematically assess an organization's abilities to achieve its digital preservation goals. A wide variety of assessment tools exist for this purpose. These range from light-weight checklists to resource-intensive certification processes. Conducted as part of the BenchmarkDP project, this paper presents a survey of these tools that elucidates available options for practitioners and opportunities for further research.

General Terms

Institutional opportunities and challenges; Frameworks for digital preservation

Keywords

capability, maturity, risk, organizational assessment, design science

1. INTRODUCTION

Over the past two decades digital preservation (DP) research has produced a wide range of tools, models, strategies, and other innovations to facilitate the long-term management of digital objects. Although much progress has been made in this area, solutions targeting individual components do not work in isolation and consideration must be given to digital preservation capabilities at the organizational level. Unfortunately, the DP community currently lacks standardized assessment tools to facilitate rigorous and systematic evaluation of an organization's capacity to achieve its preservation goals. Systematic assessment at the organizational level is essential to evaluate the efficacy of an organization's DP operations, to provide reliable benchmarks against which continuous improvement can be made, and to enable comparisons across institutions.

The BenchmarkDP project is developing and evaluating rigorous, systematic, and evidence-based means for comparing

techniques, approaches, tools, and systems. As part of the project's ongoing study of organizational assessment in DP, this paper provides a comprehensive survey of existing models and frameworks that assess an organization's ability to achieve its DP goals through a combination of people, technology, and processes. The survey is driven by two research questions: (RQ1) *What are the options for organizational assessment, and how do they vary in terms of focus, requirements, and expected outputs?* (RQ2) *What trends and gaps exist in the current landscape, and do these present opportunities for research?*

In answering these questions, the survey will aid practitioners in comparing the different options for organizational assessment, including the strengths and limitations of each approach. As well, this work will outline potential new directions for researchers and highlight areas where further study is needed. The first sections of this paper provide a brief background in assessment and key concepts, the rationale for the selection of the models surveyed, and a brief description of each model. A more detailed analysis and discussion follows. A concluding section highlights gaps in the current spectrum of solutions and identifies opportunities for further research.

2. BACKGROUND

The long-term focus of DP requires a set of coordinated activities and supporting infrastructure that includes people, technology, systems, information, and processes. This work is carried out by an organization (or an organizational unit that is part of a larger body) with the responsibility of preserving and providing access to digital information. As the field of DP matures, more systematic methods of understanding and comparing these activities are needed in order to assess the current state of preservation capabilities, identify areas that need improvement, and direct improvement efforts. Organizational assessment provides a method of measuring current performance and enables steps towards increased capacity, improved reliability, demonstrated trustworthiness, or reduced risk.

Outside of DP, this challenge of organizational assessment has been approached in different ways. We focus here on maturity models, as they are a prominent means of systematic assessment in other fields, with existing foundations to draw on. Maturity models generally, and Capability Maturity Models specifically, can be used to take an informed approach to continuous improvement[29].

iPres 2015 conference proceedings will be made available under a Creative Commons license.

With the exception of any logos, emblems, trademarks or other nominated third-party images/text, this work is available for re-use under a Creative Commons Attribution 3.0 unported license. Authorship of this work must be attributed. View a copy of this licence at <http://creativecommons.org/licenses/by/3.0/legalcode>.

These concepts originate in the Software Engineering Institute Capability Maturity Model (CMM) developed to ensure reliable and consistent processes within the field of Software Engineering [35]. While many models and frameworks have been developed based on this original CMM, there is still little consensus or consistency in the meanings and uses of the terms ‘capability’ and ‘maturity’ [25]. We define capability broadly as the sustained ability to achieve a goal, through a combination of people, technology, and process[7]. Maturity is more difficult to define. Others have noted the different senses or aspects of maturity that are often confused [25], but all stem from the common dictionary definition of ‘a state of completeness.’ What is of primary interest for assessment is the process of bringing something to maturity, the path to completeness [25]. To achieve this, maturity models describe the different sequential stages of growth – an ‘evolutionary path’ – that target individual processes or multiple dimensions. An organization’s overall state of maturity provides a measure of how much confidence one can have in the organization to successfully achieve goals and consistently provide services – in short, the degree of reliability and predictability.

Recent work demonstrates a growing interest in assessment through maturity model frameworks in other domains[50], and we see this growing interest mirrored in DP. While we include many models in our survey that are not formal CMMs and do not use the vocabulary of maturity models generally, we maintain that both of these dimensions (the capabilities available, and the predictability in successfully achieving goals) must be addressed for organizations to meet the challenges of DP. We will therefore use these concepts, and the associated literature on maturity models from other domains, to examine frameworks for organizational assessment.

In describing these existing approaches to organizational assessment in the domain of DP, we will discuss models, methods, tools, and frameworks. For our purposes, **models** are a “formal description of some aspects of the physical or social reality for the purpose of understanding and communicating” (Mylopoulos, quoted in [29]). Mettler describes **methods** as “systematic (i.e. they deliver rules on how to act and instructions on how to solve problems), goal-oriented (i.e. they stipulate standards on how to proceed or act to achieve a defined goal), and repeatable (i.e. they are intersubjectively practicable)” [29]. A **tool** is a concrete or abstract thing used to perform a task¹. Finally, a **framework** is the overall set of components, including at a minimum a model, and any associated methods or tools.

3. OVERVIEW OF FRAMEWORKS

Many assessment frameworks are referenced in discussions of DP. We cast a wide net for this survey, with literature searches in Scopus and Google Scholar for permutations of ‘digital preservation’ and ‘tool,’ ‘framework,’ ‘model,’ ‘capability,’ ‘maturity,’ ‘measurement,’ ‘improvement,’ and ‘assessment,’ as well as snowball sampling of the citations from this initial set of literature. Community venues and websites were explored, such as the Preservation and Archiving Special Interest Group (PASIG), the Digital Preservation Coalition, and the blog *The Signal*². Additional models

¹Oxford English Dictionary “tool, n.” www.oed.com

²<http://blogs.loc.gov/digitalpreservation/>

Table 1: Introducing the different Frameworks

Name	Abbreviation	Year
The Five Organizational Stages of Digital Preservation [26]	Stages	2003
Capability Assessment and Planning Toolkit [39]	CTG	2005
DRAMBORA [27]	DRAMBORA	2007
JISC AIDA Toolkit [37]	AIDA	2008
Data Seal of Approval	DSA	2010
ISO16363 / TRAC [4]	ISO16363	2012
Digital Preservation Capability Maturity Model [19]	DPCMM	2012
SHAMAN capability model [6]	SHAMAN	2012
Maturity Levels [13]	Brown	2013
NDSA Levels of Digital Preservation	Levels	2013
DIN31644 / NESTOR Seal [2]	NESTOR	2013
NSLA Maturity Matrix [36]	NSLA	2013
Scoremodel	Scoremodel	2013
e-ARK [41, 40]	e-ARK	2015

were included based on our own familiarity with existing publications.

We then developed a set of inclusion and exclusion criteria. First, the assessment must be explicitly focused on the domain of DP. Many models address closely related domains such as Records Management or Information Governance. However, maturity models by JISC and ARMA as well as the ECOMM³ were excluded since they do not address concepts or concerns specific to DP such as those outlined in OAIS[3] or TDR[42]. Similarly, the CMM for Scientific Data Management [16] was excluded as it addresses concerns specific to research data, and only covers DP from a high-level perspective.

Further, within the domain of DP, we included only models that target the organization (or organizational unit). We excluded the SPOT model for Risk Assessment[49], Data Curation Profiles Toolkit⁴, and the Data Asset Framework⁵ as they target a specific function only. The PLATTER framework[18] and NEDCC checklist[12] were also excluded as they cover initial planning but not systematic assessment for improvement. The Preservica DPMM⁶ was excluded since it targets storage media, or storage services. Finally, practical criteria were considered — the model and assessment framework must be freely available online, and in English. Due to language barriers, the Dutch ED3⁷ and the German DIN31644 standard were excluded.

In total, 14 models met all criteria, described briefly below in chronological order and listed (with abbr.) in Table 1.

Five Organizational Stages of Digital Preservation (2003)

– This is the earliest example of a model for organizational assessment and improvement focused on DP. Its motivation stems from an attempt to shift discussions away from technologically oriented solutions, towards ‘or-

³https://ecmmaturity.files.wordpress.com/2009/02/ec3m-v01_0.pdf

⁴<http://datacurationprofiles.org/>

⁵<http://data-audit.eu/>

⁶<http://www.preservica.com/download/852>

⁷<http://www.den.nl/standaard/225/>

ganizational response and readiness' issues. The target audience is defined broadly as all 'cultural repositories,' but examples used draw heavily on work with research libraries (mostly the authors' experiences at Cornell University Library). A community-created model, the structure is simple and lightweight, with three key indicators (policy and planning; technological infrastructure; content and use) for each of the five stages. It uses a conception of DP based on the three-legged stool model: organizational infrastructure, technological infrastructure, and resources framework.

Center for Technology in Government (CTG) Capability Assessment and Planning Toolkit (2005) – This model was released by the CTG at SUNY Albany. Built on the basis of the UNESCO Guidelines for the Preservation of Digital Heritage[30] and the *Stages*[26], it was developed in collaboration with the Library of Congress, with input from the broader community. It is intended to guide self-assessments of the DP capabilities of state governments and government agencies, to be used by a range of librarians, archivists, records managers, and other information professionals. The assessment process, conducted through a series of workshops, aims to identify gaps and weaknesses in 19 dimensions of capability. The toolkit provides a range of useful templates and examples.

DRAMBORA (2007) – The Digital Repository Audit Method Based On Risk Assessment was created as a joint project of the Digital Curation Centre (DCC) and Digital-PreservationEurope (DPE). This model approaches organizational readiness through risk assessment, complementary to other risk models that are not the focus of this survey [49, 10]. The 'internal audit' method progresses through 6 stages, beginning with documenting contextual information about the organization then identifying specific assets and activities, and risks, with probability and impact. A common framework of four Operational functions and four Support functions is used throughout the audit. There is an offline toolkit of templates (MS Word and Excel files), and an online form that streamlines the process and provides a summary report. Use of the online toolkit further allows for comparison with results from similar repositories that have completed the assessment.

JISC Assessing Institutional Digital Assets (AIDA) Toolkit (2008) – The AIDA Toolkit, created by the University of London Computer Centre, operationalizes the Stages through a self-assessment tool aimed at evaluating the digital asset management and DP readiness of higher education institutions in the UK [37]. The Toolkit Mark I was released in 2008 and a revised Mark II released in 2009. The objective of this toolkit is to capture an accurate picture of the current state of an organization's readiness and capabilities, not to provide explicit direction for improvement. The assessment process is based on a weighted score system that the AIDA project team requests in order to provide feedback. The toolkit contains templates and scorecards as well as an online tool.

Data Seal of Approval (DSA) (2010) – The DSA⁸ is a simple list of criteria, and an online tool, created by Dutch-based DANS. It consists of 16 guidelines, in 3 categories:

Related to Data Producers, Related to Repositories, and Related to Data Consumers. Each guideline receives a rank of 0-4 based on the five possible responses/statements, the highest being 'implemented.' The assessment is presented as a two-tiered model, supporting self-assessment for internal improvement and a 'seal of approval' for meeting the guidelines, verified through a peer-review process. Between the initial release in 2010, and the current version from 2013, 41 seals have been awarded. All material for conducting the assessment is freely available online, including manuals for applicants and reviewers. All of the documentation from the awarded seals is available on the DSA website.

ISO16363/TRAC (2012) – ISO16363:2012[4] is a standard for an audit process of the trustworthiness of digital repositories, based on compliance with ISO14721 (OAIS)[3]. It builds on the influential 2002 report which outlined the attributes and responsibilities of a Trusted Digital Repository[42] and the subsequent and the subsequent collaborative work between RLG, NARA, and CRL which resulted in the Trustworthy Repository Audit and Certification (TRAC) Criteria and Checklist[34] published in 2007. The audit process for the standard is lengthy and resource-intensive. It takes into account a wide range of organizational, infrastructure, security, and management factors. Certification is available through organizations such as the Centre for Research Libraries (CRL) and the Primary Trustworthy Digital Repository Authorization Body (PTAB), usually at considerable cost. Several repositories have been certified using this process, and there is some indication that the standard can be used for self-assessment. A tool⁹ developed by MIT has been built on this standard. A standard outlining requirements for bodies providing certification based on ISO16363, has recently been released as ISO16919:2014.

DPCMM - Dollar & Ashley (2012) – This DP Capability Maturity Model was created by consultants Charles Dollar and Lori Ashley. The model was first made available in 2012, with an updated version released in 2014. Based loosely on other CMMs, the model defines five levels or stages of capability in digital preservation: Nominal, Minimal, Intermediate, Advanced, and Optimal. The requirements for each level are specified for 15 different components, covering both Infrastructure and Services. The model is largely based on OAIS, drawing heavily on the model's concepts and vocabulary from these standards and using compliance with ISO14721 as a threshold for certain stages. For each component a table is presented defining requirements to achieve each level, paired with a score from 0-4. Scores are then summed to provide the Aggregated Digital Preservation Capability Index Score. Additional reports, such as a 'road map,' are understood to be provided if the assessment is undertaken by the consultants. They have also introduced an online tool, available at www.digitalok.org.

SHAMAN capability model (2012) – The SHAMAN Reference Architecture[6], based on enterprise architecture concepts, includes a capability-based model of DP that identifies 11 capabilities in three groups: governance, risk, and compliance capabilities; business capabilities; and, support capabilities. The emphasis for preservation is on the four capabilities (acquire content, preserve bit streams, preserve

⁸<http://datasealofapproval.org/en/>

⁹www.archivematica.org/wiki/Internal_audit_tool

content, and disseminate content) that comprise the category of business capabilities and which are supported by the capabilities in the remaining categories. This capability model was further developed into a checklist assessment method that contains five steps: identify stakeholders, identify influencers, derive preservation goals, determine capabilities, assess capability level [8].

Adrian Brown’s Maturity Levels (2013) – Presented in 2011 and subsequently published in a book [13], the development of this model was inspired by P2MM from the field of project management. No specific methodology is described for the assessment. Instead, each process perspective and capability level is paired with a statement about an action taken or process in place, such as “A written, approved digital preservation policy exists.” No specific statements are provided for the lower levels of ‘awareness,’ so the model only provides a three-level scale of Basic Process, Managed Process, and Optimized Process. We have not encountered any applications of this model.

NDSA Levels of Digital Preservation (2013) – This model is a tiered matrix of practical recommendations created by the National Digital Stewardship Alliance.¹⁰ Intended to assist organizations in establishing and improving DP activities, this model can also be used to assess the level of preservation achieved for specific digital objects or groups of digital objects. It was intended to achieve a middle ground between the complexity of ISO 16363 and simple guidance checklists [22]. The *Levels* consist of five dimensions and four progressive levels of maturity. With a focus on five functional areas at the operational level, the model is missing many of the organizational elements or dimensions present in other frameworks, but is nevertheless useful for certain applications.

DIN31644 / NESTOR Seal (2013) – Based on the NESTOR Catalogue of Criteria for Trusted Digital Repositories (2006)[32], the NESTOR Seal is an extended self-assessment process for digital archives, covering 34 criteria separated into three areas (organizational, technical, and infrastructure & security). It is based on the German DIN31644 ‘Criteria for trustworthy digital archives,’ but structured in a way similar to the DSA, providing an extended reviewed self-assessment. An organization may apply for a NESTOR Seal that recognizes compliance with these criteria (but is not an accredited certification) by providing the documentation of their self-assessment for review by NESTOR. The review will be completed within three months; there is a 500 Euro fee for applying for the seal. It is envisioned as the middle-ground between the lightweight assessment of the DSA, and the intensive auditing of ISO16363. The full text of the DIN standard is available only in German, but the criteria have been summarized in English for the NESTOR seal in an ‘Explanatory Notes’ document [33].

NSLA Digital Preservation Environment Maturity Matrix (2013) – This model was created by the National & State Libraries of Australasia DP Group. Based on OAIS, this work aims to determine digital preservation maturity in relation to the OAIS Functional Entities through a five level CMM derived from the original SEI CMM[35]. Each Func-

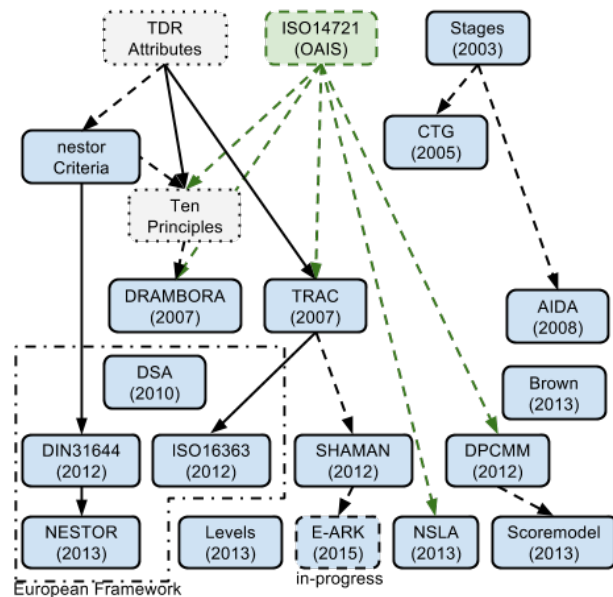


Figure 1: Family Tree of the different Frameworks.

tional Entity is associated with one of the five levels during the assessment process through a series of questions completed by the auditor. The purpose is to identify the levels of maturity, development needs, and collaboration needs of the NSLA member institutions. No recommendations or guidelines for improvement are provided by the Maturity Matrix, however it includes assessment templates.

Scoremodel (2013) – Scoremodel¹¹ is an online tool to identify risks and threats to digital objects, as well as provide basic recommendations. It is organized around seven clusters: organization and policy, preservation strategy, expertise and organization, storage management, ingest, planning and control, and access. In each of these sections, the tool presents users with a series of yes or no questions, each with context, associated risk and risk level, and an example of the evidence to be considered in the answer. Scoremodel is free to use online, open to all users, and available in both English and Dutch. However, the concepts, model, and rationale behind the tool are unclear and limited documentation is available.

E-ARK (2015) – In early 2015, the European Archival Records and Knowledge Preservation project released a maturity model for information governance that included many components outside the boundaries of this study. This work has continued with the release of an archiving maturity model, including an initial assessment and evaluation of a pilot study of 7 Archives released in October 2015. This model draws on TRAC and OAIS and presents a self-assessment questionnaire of 35 questions with responses corresponding to five levels of maturity. Questions are also grouped into five main capability areas: Pre-Ingest, Ingest, Archival Storage and Preservation, Data Management, and Access.

The relationships and influences of these models are mapped in a ‘Family Tree’ shown in Figure 1. This diagram also includes other influential documents and standards (indicated

¹⁰<http://digitalpreservation.gov/ndsa/activities/levels.html>

¹¹<http://scoremodel.org/en>

by dashed boxes) that are not assessment models, specifically the OCLC/RLG attributes for Trusted Digital Repositories (TDR)[42], ISO14721/OAIS (2003, revised 2012), and the Ten Principles, developed jointly by CRL, DCC, DPE, and NESTOR (2007)¹². Solid arrows show direct connections and evolution of models, while dashed arrows indicate explicit but loose influence.

4. ANALYSIS

4.1 Analytical Framework

An analytical framework is necessary to better understand this wide-ranging set of models, to find insightful patterns or trends. There is a growing body of literature studying maturity models in other domains, and we draw from this work in our analysis. In particular, the broad field of Information Systems has developed a rich body of knowledge on Design Science research methodology that approaches problem-solving through study and iteration of designed artifacts [23], and this previous work includes development of design principles in relation to maturity models and assessment frameworks.

Wendler[50] notes the variety of research that exists on maturity models, and we have attempted to cover a wide range to form our theoretical foundation here. First, the work by Jokela et al.[24] provide a similar survey of models in the domain of usability that focuses on the application of models. Second, to understand the models as artifacts, we have drawn on work in Design Science research, including examples and approaches define requirements for the process of developing a model [11], as well as general design principles for maturity models [38].

We determined a number of attributes to address our first research question regarding the existing options available for organizational assessment (and their focus, requirements, expected outputs). We first determined the **primary purpose**, understood here as the intended central aim of the model, and the motivation for undertaking the assessment. We defined three categories for primary purpose: initial planning, improvement, and certification. We also examined the nature of the assessment process and expected outputs. Specific requirements are necessary for different types of intended **audience** of the model, e.g. to be shared internally in the organization, with external stakeholders, or both. As well, we considered the **mode of application** (or ‘method of application’ in [17]), e.g. whether it is performed as a self-assessment, third-party assisted, or by a certified practitioner.

Next, we examined the degree of **concrete guidance** provided by each model, understood as the amount of clarity and documentation provided for applying the model, and the method of assessment [24]. This should also reflect that the method is ‘systematic, goal-oriented, and repeatable’ [29]. Here we extend this to include a discussion of the degree of detail or granularity provided in the results of the assessment and recommendations for improvement. We further noted which models provide formal documentation of **methods of assessment**, and what other **tools** are available for use. (Methods and tools are defined in Section 2 above). Additionally, **Empirical evidence** is used to describe if and how

use of a model is substantiated [24]; we have used a broad definition here to consider evidence of any/all applications, including case studies.

Finally, we note that Wendler[50] draws a distinction between research conducted with maturity models and research conducted on or about maturity models. Research ‘with’ maturity models includes all research related to the development, application, and validation of a model. Research ‘on or about’ maturity models can be seen as the “meta” work that takes the maturity models themselves as the subject of research. One of the salient conclusions of Wendler’s mapping study was that there was a need for further research ‘on or about’ models, and that the development of such work can have significant implications for both researchers and practitioners. Research on or about models would lead to fewer, but better (theoretically rigorous and empirically validated) models, discussed further in Section 5.

4.2 Analysis and Results

The analytical framework reveals a number of patterns and common traits of these models, and the overall results of the analysis are summarized in Table 2. Examining the intended **audience** and **purpose** of the models reveals that almost all ‘planning’ and ‘improvement’ models are intended primarily for internal audiences. Only three certification-oriented models (DSA, NESTOR, ISO16363) were found, and are all part of the European Framework for Audit and Certification of Digital Repositories¹³ that reflects a path of progressively rigorous audits. While intended for external audiences, assessments with ‘certification’ models can also be used internally.

The majority of the models use self-assessment as the **mode of application**, though some pair self-assessment with third-party assistance, such as the peer-review methods of DSA, NESTOR, AIDA and NSLA. DPCMM is the only model using third-party assistance through a commercial consulting service. Brown and e-ARK provide no clear documentation of application, and are noted as N/A. ISO16363 is the only model intended for assessment by a certified practitioner, to be standardized through ISO16919 “Requirements for bodies providing audit and certification of candidate trustworthy digital repositories.” [5]

A key finding of this analysis is that most models provide little **concrete guidance** for assessment or subsequent improvement measures. The ‘certification’ models provide more thorough documentation and the ‘initial planning’ models provide the least. A handful of models provide documents describing **methods of application** (DRAMBORA, SHAMAN, DSA, NESTOR). However, most provide little direction or instruction for application or result in recommendations for improvement (though commercial products like DPCMM may have a more detailed method of assessment that is not publicly available). Nine of the fourteen models provide some kind of **tool** to aid in carrying out the assessment. These range from paper-based templates to electronic forms to interactive online tools. Some online tools can be seen as providing an implicit step-wise method, however, this is not made explicit.

¹²<http://tinyurl.com/qgnt367>

¹³<http://www.trusteddigitalrepository.eu/Memorandum%20of%20Understanding.html>

Table 2: Surveyed models, methods and tools for organizational assessment in DP.

Name	Audience	Method	Tool	Mode	Concrete Guidance	Empirical Evidence
Primary Purpose: Initial Planning						
Stages	internal	-	-	SA	Limited. Key indicators note high-level processes	Examples from Cornell, no further case studies.
Levels	internal	-	-	SA	Set of practical recommendations for use exists.	Content-based case studies
Primary Purpose: Improvement						
CTG	both	-	PT	SA	Significant guidance for application through workshops, including template usage and data gathering.	Extent of use is unclear, limited evidence.
AIDA	internal	-	EF/ OT	SA (TPA)	Limited to instructions for tool; recommendations and feedback provided by project team	Multiple applications, but little documentation or evidence available
DRAM-BORA	internal	Y	EF/ OT	SA	Guidance documents are available, very detailed results	Extensive and well documented
Brown	internal	-	-	N/A	Limited. High-level processes identified	None
SHAMAN	external	Y	-	SA	No guidance on using the model	None
DPCMM	internal	-	OT	TPA	Limited to the description of the model	Model has been applied, but no documentation or evidence
NSLA	both	-	PT	SA (TPA)	Some guidance is provided for use of the tool. Results are limited to identifying areas of weakness	Only the initial study for which the tool was created
Score-model	internal	-	OT	SA	Limited recommendations both for use and in results	Some previous assessments can be seen.
e-ARK	internal	-	OT	SA	Limited to description of model	Results of pilot study available
Primary Purpose: Certification						
DSA	both	Y	OT	SA, TPA	Guidance documents are available for applying for seal	Many applications, publicly available documentation, some published case studies
ISO16363	both	-	(OT)	all	Guidance documents are available for conducting audit	Many applications and case studies
NESTOR	both	Y	EF	SA, TPA	Little guidance beyond addressing documentation to provide for seal	Multiple applications, but limited evidence

Legend: PT= paper templates; EF=electronic forms; OT=online tools;

SA=Self-assessment; TPA=Third-party assisted; CP=Certified Practitioner; all = SA, TPA and CP

Further, many of the models are supported by no **empirical evidence** at all, with only very weak indicators of successful application (such as case studies), and no direct supporting evidence. The ‘certification’ models provide the greatest number of examples in terms of application documentation and case studies, but still provide little empirical evidence to establish user trust or demonstrate validity.

5. DISCUSSION

5.1 Trends and Tensions

This analysis of models and frameworks has generated a number of insights into both the larger field of DP and the models themselves, as well as shedding light on the tensions around systematic assessments modeled after CMMs.

One significant trend to emerge from this comparison is a marked increase in the number and complexity of models in recent years. This increasing interest in assessment models mirrors the increasing number of operational repositories and commercial offerings [46] and corresponds to the findings in a recent survey [14]. However, a greater number of models has not helped to address the challenges associated with assessment, and it is increasingly difficult to weigh the costs and benefits of different approaches. There are still tensions between standards-compliance and improvement, and balancing simplicity in carrying out the assessment with reliability or trustworthiness of the results. Reliability of re-

sults is often only achieved with significant investments of time, effort and cost. Even then, few models currently provide results that can be used to directly inform planning or decision-making; others note that the existence of such a decision mechanism for improvement paths is a fundamental design principle for prescriptive use of maturity models [38].

Many models use a numeric rating, and also translate each level into direct questions of which criteria are met. However, experience in process assessment has shown that translating criteria into questions does not result in accurate descriptive results. ‘If you want to assess the maturity of a process, you do not take the direct approach of asking people whether they think the ... process is managed or established in their organization.’[9]. Similarly, challenges in finding consensus on ratings using direct questions are unsurprising. In fact, the SEI Appraisal Requirements for CMMI forbid the usage of numeric ratings if the assessment does not meet the stringent requirements of the highest-class assessment method[48].

Increased interest and development of assessment models can indicate the field’s transition from a ‘skilled artisan’ orientation towards the emergence of industrialization and professionalization, as described by McKinney[28], though this shift is not always beneficial or desired. Assessment frameworks come with assumptions that sometimes conflict with

the reality in many DP situations. Improvement is often oriented towards quality control and consistency, minimizing variability of outcomes over time and reducing individual agency. Culture built around the work of skilled artisans can contrast sharply with these assumptions, resulting in resistance to the transition to an industrial era [28].

The assumptions of sophisticated organizational assessment frameworks such as those compliant with ISO15504[1], a standard for process assessment in software development partially derived from CMM by the SPICE (Software Process Improvement and Capability dEtermination) Working Group[20], include a process orientation, the availability of multiple instances of the assessed processes across the organization's resources, and a depth and distribution of knowledge. These cannot always be assumed. Just as the CMM was not universally praised in the software industry [21], current highly detailed standards prescribing functional requirements for repositories are not necessarily fit for all purposes. Additional tensions of using models that are reflective of 'industrial era' thinking include the tendency to oversimplify reality through CMMs and the obscuring of alternate paths to maturity [44].

The frameworks surveyed here that do explicitly draw on existing CMMs do not distinguish between capability and maturity, project and process, compliance and improvement. Where they do declare adherence to a model such as the SEI CMM, they often do not demonstrate awareness of the concepts and assumptions. In general, greater clarity about underlying concepts and a stronger adherence to design principles for maturity models is needed to instill trust in these frameworks.

Finally, we note that while the CMM approach provides a framework for systematic assessment it focuses on a sequence of events or activities, not on influential factors[44]. Since CMMs were created initially to address process improvement in large organizations devoted to engineering this work is understood to be project-based, and focused on product development. None of these assumptions hold for a typical organization in the domain of DP. DP is often undertaken in small organizations or organizational units, and is not a project-based endeavour resulting in an end product that can be tested for quality and consistency. Some of these assumptions have been dropped in subsequent developments such as CMMI-SVC[47] focused on service delivery; however, current reference models in DP are not based on the principles of service-orientation. Therefore, while addressing capabilities and processes is useful, we may also need to consider the impact of other influential factors over the long-term timeframes necessary for digital preservation.

5.2 Implications for Practitioners

Generally the models surveyed provide limited guidance for conducting an assessment. Together with the absence of empirical evidence, which leads to a lack of trust in the diagnosis, this can present problems for practitioners. Below, we discuss these implications, as well as requirements and expected results, grouping the models by primary purpose.

Certification – There is a clear, but narrow, set of choices for certification: DSA, NESTOR, and ISO16363. These are generally resource-intensive, and make heavy demands on

documentation, time, and effort.

The DSA has the least stringent requirements. The process consists of a self-assessment conducted with the online tool and submitted online for review. Required time and resources largely depend on the availability of documentation within the organization. The full self-assessment can take as little as four person days to complete.¹⁴ No site visit from an auditor is required, and the peer review process conducted by the DSA takes approximately two months. Referenced documentation must be made available online for certification. Once granted the seal will need to be updated periodically as the terms of compliance change.

Certification with the NESTOR Seal is similar to the DSA, but has greater demands. NESTOR requires two contact people at the organization to assume responsibility for correspondence during the two-stage review process of the self-assessment and supporting documentation which takes approximately three months.

ISO16363 demands the most of organizations to complete an assessment. The process of certification requires extensive preparation including a thorough self-assessment against the Standard's 84 criteria and the preparation of a full catalog of relevant documentation. This option requires site visits from auditors, who themselves must meet the requirements outlined in ISO16919, and has been shown to take at least six months for many organizations [15].

Choosing from these three options will depend on the particular circumstance of an organization including the availability of documentation, willingness to commit time and resources to the assessment process, and the perceived benefits of certification in relation to the organization's objectives. All three of these assessment frameworks assume a certain degree of maturity and are not oriented towards planning for improvement, but towards compliance with ISO14721 (OAS). There is potential for a mismatch, if the organization has not adopted the OAS Reference Model, as David Rosenthal has noted from his experience.¹⁵

Improvement – As the analysis demonstrated, assessments for improvement vary widely. Decisions in this area can be structured by three factors: (1) the need for tools to conduct the assessment, (2) the major concerns practitioners wish to address, and (3) the availability of third-party assistance.

Nine of the improvement frameworks include various types of tools. Of these, two provide paper templates (CTG and NSLA), three provide both electronic forms and online tools (AIDA and DRAMBORA), and three provide stand-alone online tools (DPCMM, Scoremodel, DSA). Attempting to use these models without reference to the tools provided may compromise the results of the assessment. Third-party online tools may also be available (e.g. for ISO16363).

Organizations seeking third-party assistance in conducting an assessment for improvement have few options. AIDA,

¹⁴see the Archaeology Data Service (ADS) case <http://www.dcc.ac.uk/resources/case-studies/ads-dsa>

¹⁵<http://blog.dshr.org/2014/08/trac-audit-lessons.html>

CTG, and DPCMM offer different degrees assistance for use of their models. AIDA and CTG are the products of projects and the extent of support and future availability are unknown. Both projects were completed more than five years ago and have shown limited activity in recent years. DPCMM is active through the consulting services, however practitioners should be aware that assessment as a commercial service may have implications for the trustworthiness and reliability of the results. The NSLA model was developed for both internal and external assessment, however it is not clear that third party assistance was ever offered to organizations outside of the consortium.

When selecting a model for organizational assessment, practitioners should be mindful of the fact that with few exceptions, the models for improvement suffer from little or poor documentation, unclear theoretical foundations, and limited transparency. As such, non-certification models raise concerns about reliability and general applicability. All of the improvement models vary on these points, but none are as rigorous as the certification processes.

Initial Planning – Practitioners looking for assessment for initial planning have two options (*Stages* and *Levels*). These models use self-assessment to produce outputs targeted at an internal audience, with less focus on ongoing improvement.

Gaps: Requirements and Outputs – Tensions exist regarding the requirements for, and outputs of, the organizational assessment models currently available in DP. Organizations require well-grounded and robust assessment models with clear methods that produce reliable outputs. The few models that provide full-fledged methods, trustworthy outputs, and meaningful scores also place heavy demands on time and resources that few organizations can afford.

The gap between requirements for organizational assessment in DP and the current range of available options, is particularly significant in light of the degree of development of maturity models in other fields. More sophisticated assessment methods, such as those compliant with ISO15504, make assumptions about process-orientation that do not hold true for many digital repositories seeking assessment. More problematic still, is that many of these models provide ratings that provide the impression of comparability, but without this solid basis. Those gaps point to manifold opportunities for research.

5.3 Implications for Researchers

This survey demonstrates a need for further study of the various types of models available for organizational assessment within DP. First, the work begun here can be extended, and more detailed evaluations of specific models should be completed using principles of maturity models and design science research. Additionally, we have identified the need for further research ‘with’ models, separated into three areas: development, application, and evaluation of models.

Research Developing Models – Future research can expand on the concept of frameworks and approaches that form the basis of the different models. While the majority of models use a framework related to concepts of capability and/or maturity, few (if any) provide a full definition of these concepts, or demonstrate how they have drawn on

the existing research in this area. Shared frameworks provide the benefit of a cumulative tradition, with new work building off the foundations of previous model development. There may be other types of frameworks beyond maturity models that are useful, and an argument can also be made for more diversity in the frameworks used. Additionally, there is a lack of theoretical grounding, or direct evidence of this grounding, in the models studied, particularly around the development of models as designed artifacts. Further work is necessary to determine when and how design principles or guidelines, as described in Design Science research on maturity models, are evident in different models.

Research Applying Models – Research on the application of models is currently limited by the lack of documentation and evidence. Many frameworks do not specify methods (with the exception of DRAMBORA, DSA, and NESTOR), and there is generally little concrete guidance or documentation on carrying out assessments. This is an essential missing component; a robust assessment framework must consist of both a model and a method for its application in order to ensure that assessments are systematic and repeatable. We found that a limited number of case studies exist that describe the details of the application of the model in practice, and these were only available for a handful of models (DSA, ISO16363, and DRAMBORA). This is an area that can and should be explored in greater detail, and in particular there is a need for more rigorous case studies carried out by researchers not associated with development of the model. This is reflected in the recent NDSA Agenda[31] that emphasizes the need for greater large-scale evidence-sharing and capacity-building in the DP community. As noted above, the lack of research on application methods also has implications for practitioners, who may find difficulty in applying the models without clarity of documented methods.

Research Evaluating Models – We found through this survey that, in general, more empirical evidence is needed not only to document applications and report their results, but in order to evaluate and validate the models. While there are case studies available for some models, we did not find that any focus on testing or evaluating the model itself. This is essential to ensure trust in the effectiveness of the assessment framework as an overall tool for improvement. Further evaluation might include more longitudinal studies to identify the critical success factors [43, 45] for DP. As well, engaging evaluation through a Design Science research framework will allow the results to inform the continued iterations and future development of assessment frameworks, models, and methods.

Gaps: Research ‘on or about’ models – Building on Wender’s distinction of ‘research with’ maturity models and ‘research on or about’ maturity models noted in Section 4.1, this survey demonstrates a gap, and need for more ‘research on or about’ models. This ‘meta’ approach will benefit the community as it continues to mature.

There is generally limited literature on organizational assessment in DP, and it largely, if not entirely, falls into the category of ‘research with models.’ Even then, most existing material on organizational assessment in the domain of

DP consists of papers that describe the models, their components and creation. As noted above, more work is needed that studies the development, application and evaluation of models. We propose that drawing connections with Design Science Research can benefit this work, and the field as a whole, by providing a framework to tie all these aspects together, and result in improved models as artifacts for use by organizations.

The lack of work on or about models is not unique to the field of DP, and Wendler concludes that further work is needed to address research on or about models[50]. As others are beginning to address this gap in maturity model research, we can both draw on and align with recent work from other domains. Contributing to this under-represented area will provide the community with a more solid theoretical foundation which will result in better models that are easier to use, more reliable, and more trustworthy. Future work can include concepts and theoretical grounding, definitions and dimensions of maturity, and creation of domain-specific procedures and requirements for maturity model development (such as [11]). We see this paper as a significant contribution, and a starting point for future work in this direction. To the best of our knowledge, this paper is the only contribution to the “meta” field of research on or about maturity models in DP.

5.4 Limitations

We have chosen to undertake a qualitative survey, as opposed to a structured systematic review of literature. It remains an interpretive overview, that has allowed us to characterize many models more generally. Future work could provide a more detailed review of selected models.

6. CONCLUSION & OUTLOOK

This survey attempts to make sense of the diverse and growing landscape of models and tools for organizational assessment and improvement. We have described the options for practitioners seeking to undertake an assessment, and identified trends and gaps for researchers intending to pursue further study of organizational assessment and improvement. Our analysis draws on existing work with maturity models in other domains.

We categorized the models by primary purposes of initial planning, improvement, and certification, and then outlined other requirements for their use. Many options occupy the middle ground between initial planning and more formal certification processes. These range from simple grids to more extensive documents, complete with supporting templates and tools. However, due to lack of empirical evidence, it is still difficult to estimate time and resources required for many of these assessment frameworks, as well as the effectiveness of assessment reports, results and overall organizational outcomes.

Additionally, we have argued here that the concepts of design science research, a growing/emerging approach in information systems, provides effective frameworks for future research and evaluation of models. Future work can benefit from design science principles and guidelines for development of maturity models, that can be adapted to the needs of DP. Design science also, importantly, connects development, application, and evaluation as a cycle, so that applica-

tion and evaluation continue to inform future development iterations.

We have concluded that further in-depth research and case study evaluation is needed in order to better understand the strengths, weaknesses, and appropriateness of these tools for assessing organizational capabilities. Partnerships and feedback from the community will be essential to undertake this work. We hope to continue this discussion at iPres, to better understand the tensions, needs, and potential synergies of ongoing efforts in the digital preservation community.

Acknowledgements

Part of this work was supported by the Vienna Science and Technology Fund (WWTF) through the project BenchmarkDP (ICT12-046).

7. REFERENCES

- [1] ISO 15504:2004 Information technology - Process assessment, 2004.
- [2] DIN 31644 Information and documentation - Criteria for trustworthy digital archives, Apr. 2012.
- [3] ISO 14721:2012 Space data and information transfer systems - Open archival information system (OAIS) - Reference model, 2012.
- [4] ISO 16363:2012 Space data and information transfer systems - Audit and certification of trustworthy digital repositories, 2012.
- [5] ISO 16919:2014 Space data and information transfer systems - Requirements for bodies providing audit and certification of candidate trustworthy digital repositories, 2014.
- [6] G. Antunes, J. Barateiro, C. Becker, J. Borbinha, D. Proença, and R. Vieira. Project Deliverable SHAMAN Reference Architecture v3.0, 2012.
- [7] G. Antunes, J. Barateiro, C. Becker, J. Borbinha, and R. Vieira. Modeling Contextual Concerns in Enterprise Architecture. In *EDOC 2011*, Helsinki, Finland, Sept. 2011.
- [8] G. Antunes, D. Proença, J. Barateiro, and C. Becker. Assessing Digital Preservation Capabilities Using a Checklist Assessment Method. In *iPres 2012*, Toronto, ON, Canada, Oct. 2012.
- [9] B. Barafort, V. Betry, S. Cortina, M. Picard, M. St Jean, A. Renault, O. Valdés, and P. Tudor. ITSM process assessment supporting ITIL: Using TIPA to Assess and Improve your Processes with ISO15504 and prepare for ISO20000 Certification vol. 217. *Zaltbommel, Netherlands: Van Haren*, 2009.
- [10] J. Barateiro, G. Antunes, F. Freitas, and J. Borbinha. Designing Digital Preservation Solutions: A Risk Management-Based Approach. *IJDC*, 5(1):4–17, 2010.
- [11] J. Becker, R. Knackstedt, and J. Pöppelbuß. Developing Maturity Models for IT Management - A Procedure Model and its Application. *BISE*, 1(3):213–222, June 2009.
- [12] L. Bishoff and E. Rhodes. Planning for Digital Preservation: A Self-Assessment Tool, 2007. NEDCC.
- [13] A. Brown. *Practical digital preservation*. Facet Pub., London, 2013.
- [14] E. Cardoso. Preliminary results of the survey on Capability Assessment and Improvement. In *iPres 2013*, Lisbon, Portugal, Sept. 2013.

- [15] Center for Research Libraries. CRL Report on Scholars Portal Audit, Feb. 2013.
- [16] K. Crowston and J. Qin. A capability maturity model for scientific data management: Evidence from the literature. *ASIST*, 48(1):1–9, Jan. 2011.
- [17] T. De Bruin, R. Freeze, U. Kaulkarni, and M. Rosemann. Understanding the Main Phases of Developing a Maturity Assessment Model. In B. Campbell, J. Underwood, and D. Bunker, editors, *Faculty of Science and Technology*, pages 8–19, CD-ROM, 2005. AIS, Australasian Chapter.
- [18] DigitalPreservationEurope (DPE). Repository Planning Checklist and Guidance DPE-D3.2, Mar. 2008.
- [19] C. Dollar and L. Ashley. Digital Preservation Capability Maturity Model (DPCMM) Background and Performance Metrics Version 2.6, May 2014.
- [20] A. Dorling. SPICE: Software process improvement and capability Determination. *Information and Software Technology*, 35(6-7):404–406, June 1993.
- [21] M. E. Fayad and M. Laitnen. Process assessment considered wasteful. *Comm. of the ACM*, 40(11):125–128, 1997.
- [22] A. Goethals. An Example Self-Assessment Using the NDSA Levels of Digital Preservation. In *iPres 2013*, Lisbon, Portugal, Sept. 2013.
- [23] A. R. Hevner, S. T. March, J. Park, and S. Ram. Design Science in Information Systems Research. *MIS Quarterly*, 28(1):75–105, Mar. 2004.
- [24] T. Jokela, M. Siponen, N. Hirasawa, and J. Earthy. A survey of usability capability maturity models: implications for practice and research. *Behaviour & Information Technology*, 25(3):263–282, May 2006.
- [25] A. M. Maier. Assessing Organizational Capabilities: Reviewing and Guiding the Development of Maturity Grids. *IEEE Transactions on Engineering Management*, 59(1):138–159, Feb. 2012.
- [26] N. Y. McGovern and A. R. Kenney. The Five Organizational Stages of Digital Preservation. In P. Hodges, M. Bonn, M. Sandler, and J. P. Wilkin, editors, *Digital Libraries: A Vision for the 21st Century*. Michigan Publishing, University of Michigan Library, 2003.
- [27] A. McHugh, R. Ruusalepp, H. Hofman, et al. Digital Repository Audit Method Based on Risk Assessment (DRAMBORA), 2007. eprints.ermanet.org/122/.
- [28] P. McKinney. From Hobbyist to Industrialist. Challenging the DP Community, Oct. 2012. Open Research Challenges Workshop at *iPres 2012*, digitalpreservationchallenges.wordpress.com.
- [29] T. Mettler and P. Rohner. Situational Maturity Models as Instrumental Artifacts for Organizational Design. In *DESIRIST '09*, Malvern, PA, May 2009.
- [30] National Library of Australia. UNESCO Guidelines for the preservation of digital heritage, 2003.
- [31] NDSA. 2015 National Agenda for Digital Stewardship, Sept. 2014.
- [32] NESTOR Working Group. Catalogue of Criteria for Trusted Digital Repositories, Dec. 2006.
- [33] NESTOR Working Group. Explanatory notes on the nestor Seal for Trustworthy Digital Archives, 2013.
- [34] OCLC, CRL. Trustworthy Repositories Audit & Certification (TRAC): Criteria and Checklist, Feb. 2007. Ver 1.0.
- [35] M. C. Paulk, B. Curtis, M. B. Chrissis, and C. V. Weber. Capability Maturity Model for Software, Version 1.1, Feb. 1993. CMU/SEI-93-TR-024.
- [36] D. Pearson and L. Coufal. Digital Preservation Environment Maturity Matrix, Nov. 2013. National and State Libraries of Australasia, www.nsla.org.au.
- [37] E. Pinsent. The AIDA self-assessment toolkit Mark II, Feb. 2009. ULCC, <http://aida.jiscinvolve.org>.
- [38] J. Pöppelbuß and M. Röglinger. What makes a useful maturity model? A framework of general design principles for maturity models and its demonstration in business process management. In *ECIS*, 2011.
- [39] T. A. Prado, A. M. Cresswell, S. S. Dawes, B. Burke, L. Dadayan, S. Embar, and H. Kwon. Building State Government Digital Preservation Partnerships: A Capability Assessment and Planning Toolkit, Version 1.0, 2005.
- [40] D. Preença, R. Vieira, and J. Borbinha. Project Deliverable D7.2 E-ARK - Archiving Maturity Model - Initial Assessment and Evaluation, Oct. 2015. Project 620998: European Archival Records and Knowledge Preservation, www.eark-project.com.
- [41] D. Proença, R. Vieira, and J. Borbinha. Project Deliverable D7.1 E-ARK - A Maturity Model for Information Governance - Initial Version, Feb. 2015. Project 620998: European Archival Records and Knowledge Preservation, www.eark-project.com.
- [42] Research Libraries Group. Trusted Digital Repositories: Attributes and Responsibilities, May 2002.
- [43] J. F. Rockart. Chief executives define their own data needs. *Harvard Business Review*, 57(2):81–93, 1979.
- [44] M. Röglinger, J. Pöppelbuß, and J. Becker. Maturity models in business process management. *BPMJ*, 18(2):328–346, Apr. 2012.
- [45] M. Rosemann and T. De Bruin. Application of a Holistic Model for Determining BPM Maturity. *BPTrends*, Feb. 2005.
- [46] P. Sinclair, J. Duckworth, L. Jardine, A. Keen, R. Sharpe, C. Billenness, A. Farquhar, and J. Humphreys. Are you Ready? Assessing Whether Organisations are Prepared for Digital Preservation. *IJDC*, 6(1):268–281, Nov. 2011.
- [47] Software Engineering Institute, CMU. CMMI for Services, v1.3, Nov. 2010. CMU/SEI-2010-TR-034.
- [48] Software Engineering Institute, CMU. Appraisal requirements for CMMI (ARC), v1.3, 2011. CMU/SEI-2011-TR-006.
- [49] S. Vermaaten, B. Lavoie, and P. Caplan. Identifying threats to successful digital preservation: The SPOT model for risk assessment. *D-Lib*, 18(9):4, 2012.
- [50] R. Wendler. The maturity of maturity model research: A systematic mapping study. *Information and Software Technology*, 54(12):1317–1339, Dec. 2012.