

The design and use of assessment frameworks in digital curation

Christoph Becker

(corresponding author)

University of Toronto, Faculty of Information
140 St George Street
Toronto, ON M5S 3G6
Canada
phone: +1 416 9465367
Email: christoph.becker@utoronto.ca

Emily Maemura

University of Toronto, Faculty of Information
140 St George Street
Toronto, ON M5S 3G6
Canada
Email: e.maemura@mail.utoronto.ca

Nathan Moles

University of Toronto, Faculty of Information
140 St George Street
Toronto, ON M5S 3G6
Canada
Email: n.moles@mail.utoronto.ca

The design and use of assessment frameworks in digital curation

Abstract

To understand and improve their current abilities and maturity, organizations use diagnostic instruments such as maturity models and other assessment frameworks. Increasing numbers of these are being developed in digital curation. Their central role in strategic decision making raises the need to evaluate their fitness for this purpose and develop guidelines for their design and evaluation.

A comprehensive review of assessment frameworks, however, found little evidence that existing assessment frameworks have been evaluated systematically, and no methods for their evaluation.

This article proposes a new methodology for evaluating the design and use of assessment frameworks. It builds on prior research on maturity models and combines analytic and empirical evaluation methods to explain how the design of assessment frameworks influences their application in practice, and how the design process can effectively take this into account.

We present the evaluation methodology and its application to two frameworks. The evaluation results lead to guidelines for the design process of assessment frameworks in digital curation. The methodology provides insights to the designers of the evaluated frameworks that they can consider in future revisions; methodical guidance for researchers in the field; and practical insights and words of caution to organizations keen on diagnosing their abilities.

Introduction

An organization's ability to curate digital resources for extended periods is essential to enable future use and access. To fulfill their responsibilities for the provision of digital resources, organizations must understand their current capabilities – their “collective skills, abilities, and expertise” (Smallwood & Ulrich, 2004) - and their degree of maturity. To do so, they use diagnostic instruments such as maturity models and similar assessment frameworks (Maier, 2012). Maturity models consist of dimensions, criteria and scoring mechanisms that assess how an organization's capabilities improve and mature over time. These exist within assessment frameworks that often include an assessment tool or method and supporting documentation and guidelines.

A growing number of assessment frameworks aim to support digital curation practitioners in diagnosing and improving their organizations' digital curation, preservation or stewardship capabilities. These assessment frameworks structure the diagnosis and play a central role in strategic decision making. Their scope determines the focus of what is being evaluated; their criteria specify the questions asked during the assessment process and set expectations for the evidence to be examined; their thresholds define what constitutes success and achievement; and the gaps they highlight are used to prioritize investments and areas to improve (Maier, 2012). The results of such assessments provide vital input for strategic decisions about the organization's future development (Smallwood & Ulrich, 2004). This raises the need to evaluate their fitness for this purpose and to use insights from their application to support their systematic design (Wendler, 2012). However, most frameworks in digital curation do not disclose evidence about their application, evaluation and validation (Maemura et al., 2017). While the literature on maturity models more generally provides principles and requirements for the design and evaluation of assessment frameworks, it does not provide a comprehensive methodology. The question arises, *what methods can be used to evaluate assessment frameworks in digital curation, and how does the evaluation of frameworks reflect their design and use?*

To support systematic research on organizational assessment in digital curation that addresses the design, application, evaluation and evolution of maturity models, this article presents a new methodology for evaluating assessment frameworks based on design principles and requirements from previous research on maturity models. We discuss how we have implemented the methodology to evaluate two frameworks. Our approach provides an explanation of the phenomena arising in the empirical investigation of real assessment projects in two organizations. This allows us to derive lessons learned for the design process and identify priorities for future research and development.

Background

Maturity Models and Assessment Frameworks in Digital Curation

Digital curation is maturing from ad-hoc activities that address emerging stewardship concerns to a fully fledged discipline. As organizations established repository systems and organizational policies, job profiles emerged and education paths were formalized. The development of specific techniques for digital preservation and tools, models and strategies for individual aspects of digital curation raised the need to consider the relationships between these in organizational contexts. In the process of developing their abilities, organizations turn to instruments that allow them to assess where they stand and identify a path forward to improve their digital curation programs, strategies, and systems (Dobratz, Rödiger, Borghoff, Rätzke, & Schoger, 2010).

Correspondingly, the last decade has seen a steady increase in the development of frameworks that are used in practice to assess an organization's digital curation and preservation abilities (Maemura et al., 2017). Certification frameworks such as the Repositories Audit & Certification checklist (ISO, 2012) focus on compliance with strict criteria. Planning and improvement frameworks such as the DRAMBORA risk assessment method assume a diagnostic perspective that enables organizations to assess their strengths and weaknesses. We focus here on those assessment frameworks that facilitate diagnostic perspectives on an organization to enable improvement. This perspective on improvement over time, as opposed to certification, is commonly associated with the term *maturity model*. Maturity models characterize progress, explicitly or implicitly, in terms of stages and maturation paths (Pöppelbuss & Röglinger, 2011). In contrast to compliance frameworks that focus on normative criteria for how things should be, maturity models represent "theories of stage-based evolution" (Pöppelbuss & Röglinger, 2011, p. 4), and their formative assessment perspective can be very useful for decision making and prioritization of areas of improvement (Maier, 2012). In practice, various combinations of 'maturity' and 'capability' are used to denote such models and the approaches used to design and apply them, although these concepts have slightly different semantics (Maier, 2012). We use the terms *model* or *maturity model* to refer to the dimensions and criteria that focus and guide the assessment; and the terms *framework* and *assessment framework* to refer to the entire set of components within one approach, including the maturity model, an assessment tool or method, and supporting documentation and guidelines (Maemura et al., 2017).

The strategic importance of assessment results suggests that assessment frameworks should be evaluated rigorously, and that evidence about their usage should be made available to potential adopters to enable them to evaluate a framework's suitability for their context (Jokela, Siponen, Hirasawa, & Earthy, 2006; Wendler, 2012). Empirical insights into the process of diagnosing and assessing an organization are an essential element of model development (Hevner et al., 2004). Some frameworks in digital curation, such as DRAMBORA, have indeed been developed through extensive empirical work (McHugh et al., 2008). Doing so systematically requires robust methods to collect and analyze evidence of what happens in practice when these models are used.

Despite the importance of assessment for strategic planning in digital curation, however, a comprehensive review showed that very few frameworks have been empirically evaluated and documented. As a result, little is known about how different frameworks fare in supporting an organization's goals or facilitating improvement (Maemura et al., 2017). Some frameworks provide narrative reports of their application, but none provides comprehensive evidence about its characteristics and the methods used to evaluate it. Organizations often have concerns when it comes to sharing publicly the results of their assessment, since they could reveal weaknesses in their practices (McHugh, Ross, Innocenti, Ruusalepp, & Hofman, 2008). Case studies published by the creators of frameworks often focus on illustrating and advocating their use (Wendler, 2012). As a consequence, the design of maturity models in digital curation lacks robust methods that allow us to "scrutinize ... these approaches critically for the benefit of the professional community" (Maemura et al., 2017, p. 1632). This article contributes to closing this gap by establishing an evaluation method built on rigorous approaches from other disciplines. The method can be used by framework developers to examine their models, and by researchers to perform independent evaluations.

The design of maturity models

The systematic design and evaluation of such frameworks is the focus of Design Science Research (DSR), a well-established research paradigm based on the argument that designed artifacts in the context of information systems and

IT management can and should be studied rigorously (Gregor & Hevner, 2013; March & Smith, 1995). DSR emphasizes the value of research that is both relevant to practice and rigorously conducted and documented to benefit the knowledge bases of communities of researchers and practitioners (Hevner, March, Park, & Ram, 2004; March & Smith, 1995). Within this paradigm, research on maturity models focuses on their design and use for specific purposes. Principles, methods and guidelines for the design process are founded on an adherence to iterative cycles of generating design artifacts and evaluating them against specific requirements through empirical or analytical testing, following defined processes. Methodologies and procedures for DSR in information systems (Peppers, Tuunanen, Rothenberger, & Chatterjee, 2007) have found wide resonance (Gregor & Hevner, 2013).

The development of maturity models requires key design choices about a model's scope, intended audience and use, and overall structure (Mettler, Rohner, & Winter, 2010). Some models employ a grid structure that directly maps each component onto each maturity level (Maier et al., 2012). Others use complex structures to decompose maturity. For example, the Software Engineering Institute's Capability Maturity Model (CMM) defines a set of 18 process areas characterized by specific goals and activities, and five levels of maturity (Paulk, Curtis, Chrissis, & Weber, 1993). Instead of a grid, however, each stage of maturity introduces new process areas. Additionally, the CMM distinguishes between the ability to perform an activity, the commitment to perform it, and the activities actually performed. By choosing a structure to articulate maturity, a model's developers influence its possible uses. For example, the conceptual simplicity of grids makes them easier to understand and apply than the CMM (Maier, 2012).

DSR principles and methods have been used to develop specific methods, procedures and principles for the design of assessment frameworks. A set of requirements for the design process (Becker, Knackstedt, & Pöppelbuss, 2009) is summarized in Table 1.

Table 1 Requirements for the design process of maturity models (Becker et al., 2009)

R#	Description
R1	Existing models should be compared prior to development of a new model.
R2	An iterative procedure should be applied to ensure that feedback from evaluation can feed into improved versions of the model.
R3	The model should be iteratively evaluated for its "usefulness, quality and effectiveness" (J. Becker et al., 2009).
R4	Multiple perspectives should be considered in this process.
R5	The relevance of the problem that the model addresses should be clarified and explained.
R6	The problem that the model addresses should be clearly defined to allow others to evaluate the applicability of the model.
R7	The results should be presented with the target audience in mind.
R8	Models should be fully documented according to scientific standards to enable independent evaluation.

Design principles for maturity models

Throughout the design process, a set of design principles for maturity models (Pöppelbuss & Röglinger, 2011) provides an effective frame of reference for key issues to consider in the design and evaluation of such artifacts (Röglinger, Pöppelbuß, & Becker, 2012). They are grouped in three levels according to the purpose of use.

Basic Design Principles address concerns that are relevant for all maturity models.

- **Basic Information and prerequisites** includes six principles that address a model's intended use. For example, is it focused on a specific application domain, such as records management, or intended to work broadly in other domains such as research data management? The model should specify clearly the "class of entities under

investigation” and whether these are software systems, people, processes, or organizational capabilities. This crucial distinction enables the model to draw on existing concepts relevant for that class of entity, and it clarifies what statements about the entity are meaningful and what kind of evidence should be collected to support them.

- **Central constructs related to maturity and maturation** should be defined. For example, an organization may develop initial preservation capabilities by first developing staff skills and crafting a mission statement; then use this legitimization to justify the need to procure a digital repository system and initiate operational ingest. As organizations continue further on such a path, they consider aspects such as systems procurement, staff development and preservation policies (Sinclair et al., 2011). A maturity model should describe a logical path of relationships between such aspects as technical abilities, performed activities, expertise and skills, and assigned responsibilities. It will describe how these aspects and relationships evolve over time and clarify what it assumes to be the highest level of maturity.
- **Central constructs of the application domain** should be defined. For example, a maturity model for digital preservation may need to provide specific definitions for central concepts such as authenticity or integrity as the basis for defining tangible practices such as integrity checks.
- **Documentation** should be targeted and effective.

Principles for a descriptive purpose of use focus on the possibility of applying the model’s criteria consistently in an assessment project to diagnose an organization’s maturity. For example, if a criterion requires integrity checks to be carried out on a regular basis, how often is “regular”? Without such definitions, assessors in digital curation organizations will apply the criterion inconsistently and expend their energy on identifying reasonable assumptions, as reported in Miller, Blake, & Sorsby (2012). Documentation should support the assessors in adapting the model to their circumstances. For example, it might provide advice for dealing with seemingly inapplicable requirements: Should they be ignored, reinterpreted, or counted as ‘not achieved’?

Principles for a prescriptive purpose of use are relevant for models that, in addition to a descriptive diagnosis, aim to support improvement towards higher levels of maturity. The principles require models to provide guidance towards specific improvement measures. For example, such a model should support decision makers in choosing between possible improvement initiatives. Since aspiration levels vary between organizations according to pragmatic and strategic considerations, a model should provide flexible support for improvement regardless of the desired level of maturity that an organization aspires to.

The process of assessment

An assessment project is the attempt of one or more individual assessors to diagnose the current state of capability of an organization. Its process is structured by the assessment method. However, many frameworks do not specify an explicit procedure of application (Frick, Küttner, & Schubert, 2013; Maemura et al., 2017). To address this gap, we developed a general procedure for conducting the assessment in four phases (Becker et al., 2018). It is based on the steps of the CMMI appraisal method (SCAMPI Upgrade Team, 2011) but reduces the collection of ‘objective evidence’ that these auditing procedures require. Instead, the procedure focuses on self-assessment. It structures the assessment project into four phases, illustrated in Figure 1.

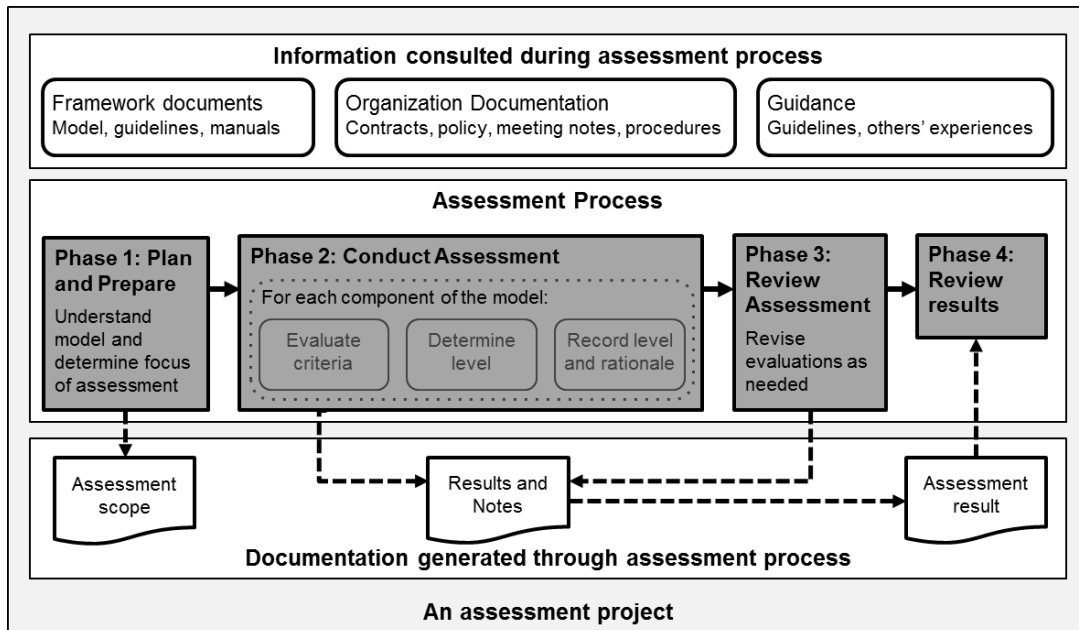


Figure 1 A basic assessment procedure to structure an assessment project

The assessment project and its outcomes thus provide the context of use in which an assessment framework can be evaluated in practice.

Research Methods

Research design and Method for Study

The outcomes of a rigorous and useful evaluation approach should provide robust insights and actionable recommendations for the design of the evaluated framework and framework design in general. The lack of such comprehensive evaluation of assessment frameworks motivates our research questions. Most centrally, we ask: *What methods can be used to evaluate assessment frameworks in digital curation? (RQ1)*. As part of this evaluation, we must examine the use of assessment frameworks in practice and make recommendations about their design, so we ask: *How does the design of an assessment framework influence its use and value in different contexts? (RQ2)*. Finally, since our research aims to support the design of future assessment frameworks, we ask: *How can the evaluation of assessment frameworks inform future design methods? (RQ3)*.

To address these questions, we have developed a comprehensive methodology for the evaluation of assessment frameworks that incorporates the DSR principles, methods and requirements summarized above. This evaluation methodology is presented in the next section. We implemented it in a study to evaluate two assessment frameworks. The Evaluation Results section presents highlights of this application of the evaluation methodology. In Research Findings, we discuss our response to each Research Question in turn.

A comprehensive evaluation methodology

The proposed methodology synthesizes the insights from the theory of maturity models within a DSR methodology that considers the disciplinary context of design and use. It incorporates three complementary perspectives of quality by evaluating the **process** of framework design, the quality of the framework as a designed **product**, and the quality of the framework **in use**. Table 2 articulates key concerns, criteria and methods for each perspective. We draw on Quality Management perspectives that have already informed digital curation management (Dobratz et al., 2010). A

distinction is made between the inter-related perspectives of process quality, product quality, and quality in use (ISO, 2001). *Process quality* refers to the design process and how it influences product quality and use. *Product quality* refers to those properties that are intrinsic to the product itself independent of a context of use (Garvin, 1984). These inherent characteristics are distinguished from the implications of a product’s use and application in real-world settings. A product’s ‘Quality in Use’ is the “degree to which [it] can be used by specific users to meet their needs to achieve specific goals with effectiveness, efficiency, freedom from risk and satisfaction in specific contexts of use” (ISO, 2011). It is influenced by product quality.

Table 2 Three perspectives of quality for the design and use of assessment frameworks

	Design process	Product quality	Quality in use
Key concern	Rigor and relevance in the process that produces the framework	Quality attributes of the resulting framework	Emergent properties of the framework embedded in a context of use
Evaluation criteria	Process requirements for the design of maturity models (Becker et al., 2009)	Design principles for maturity models (Pöppelbuss & Röglinger, 2011)	Effectiveness, efficiency, satisfaction, freedom from risk (ISO, 2011)
Evaluation method	Analytic: Review of documentation	Analytic: Study of the frameworks	Empirical: Comparative case study evaluation

The evaluation of the design process and the product demand analytical methods that examine the framework and its documentation, while the evaluation of quality in use requires empirical examination of an actual application. Therefore, our methodology combines an analytic design evaluation based on established DSR theory with an empirical evaluation.

The **design process** is evaluated analytically based on the set of requirements in Table 1. Each requirement is evaluated, cautiously relying on the available documentation to draw inferences about the design process.

The **design product** is evaluated analytically by studying the model using the set of design principles discussed above. All available documentation should be reviewed, including supporting material such as glossaries, tools, and documents describing procedures for assessment. For each principle, statements should link to evidence of adherence, and specific examples should be provided to highlight principles that are not followed.

The **quality in use** of frameworks is characterized empirically through their application in organizational contexts. The phenomenon that allows us to investigate the framework is the **assessment project** shown in Figure 1, in which a set of activities are carried out in a process to complete the assessment using the framework within an organization to produce **assessment results**: the scores, rankings, outputs, and reports that indicate the organization’s degree of maturity. Assessment projects are examined in a case study research design (Yin, 2013). This allows researchers to study real-world phenomena in depth within their distinctive situations by considering contextual factors and triangulating multiple forms of evidence. While the primary motivation of organizations to perform an assessment is to assess themselves, our evaluation focuses on what the assessment project reveals about the framework in use. The insights that emerge from observing the project and reflecting on it jointly with the assessors are captured in a case summary. The focus of examination is the **Assessment Framework** (the set of assessment models, tools and guidelines) and its quality in use.

Figure 2 illustrates the data collected to study each assessment project. Beyond the documentation generated by the project (see Figure 1), we collect data that allows us to characterize the framework in use. A pre-assessment survey establishes the assessor’s existing understanding of curation concepts and principles and their familiarity with assessment models and audits. The assessment process is recorded on video using web cameras and screen sharing. The recordings are complemented with notes by the research team and the assessor. The research team records each

component's level as well as notes and comments from the assessor. Immediately following the assessment, a semi-structured exit interview with the assessor captures their impressions of the framework, the results, and the overall process of assessment.

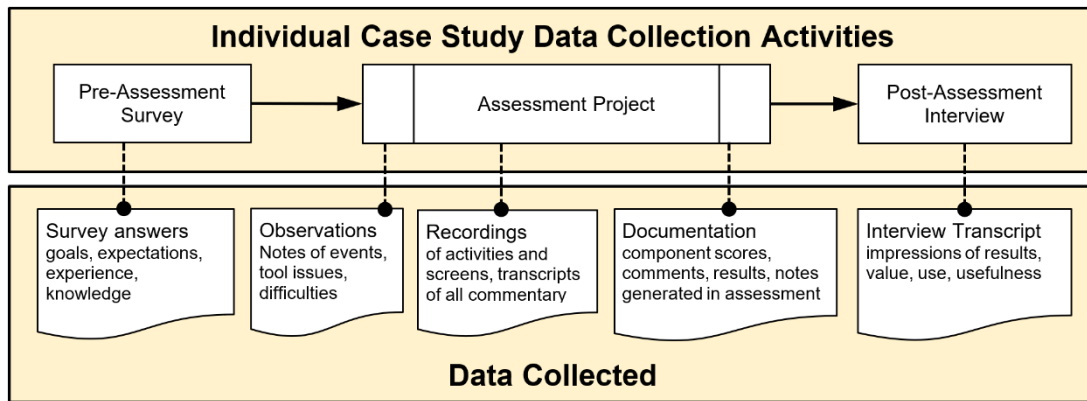


Figure 2 Individual Case Data Collection

Videos and interviews are transcribed. Qualitative coding is used to describe and summarize each case and identify aspects of interest in the observations, recordings and documents. Coding is conducted by the researchers independently and iterates until a consensus is reached.

The resulting case summary provides a succinct account of the assessment project, structured by the embedded units of analysis shown in Table 3. A detailed table captures relevant incidents such as difficulties that occurred in structured form. For each incident, it captures when (process stage) and where (model component) it occurred, what happened, and which consequences it had. This concise but detailed description of relevant events is validated by the practitioners who conducted the assessments. It provides the basis for further analysis.

Table 3 Case summaries of assessment projects should cover the following aspects.

Unit of analysis	Aspect
Assessment Process	Resources, skills, time and effort used in assessment process
	Resulting diagnosis of the organization's strengths and weaknesses
	Events observed during the process
	Incidents and difficulties
Assessment Results	Value of the results and planned use of results in the organization
	Accuracy and Reliability of results
	Assessor's confidence and trust in results

Our methodology recommends a multi-case setup. In comparisons across contexts of use, organizational differences between the assessment projects help to illuminate aspects of each framework, while different frameworks used in one organization highlight the organization's unique circumstances. This comparative analysis of multiple cases using replication logic supports theory building (Eisenhardt & Graebner, 2007) and increases the robustness of findings.

The empirical evaluation characterizes the evaluated frameworks in practice. This complements their analytic evaluation and provides an entrance point to examine the subtle relationships between process, product and use. As the findings will show, intrinsic characteristics of the frameworks that can be identified through analytic evaluation

also surface in incidents that occur in the use of frameworks in specific contexts. To explain empirically observed incidents of difficulty and derive suggestions for improving the framework, we thus connect incidents to the independent analysis of product quality. Figure 3 illustrates the resulting explanatory schema.

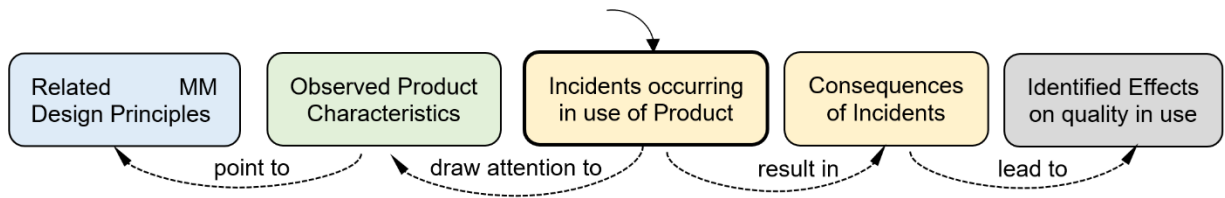


Figure 3 Conceptual links between design principles, incidents and quality in use

In the Research Findings, we discuss how this combination of analytic and empirical perspectives supports the design process.

Study Design: Implementing the evaluation methodology

We instantiated this methodology to evaluate two frameworks and their use. The empirical component of the evaluation took place in two organizations, one archive and one academic library. First, the Austrian State Archives (ASA) is the central archives of the Republic of Austria and has a legal mandate to collect, preserve, and make available all archival holdings of the federal government for research, legal and administrative purposes. Second, the University of Toronto Libraries (UTL) is a large academic library system operating as a division of the University of Toronto. Both organizations have well-established digital preservation programs, and the assessors have substantial expertise, including prior experience with ISO 16363 in a full audit (UTL) and a self-assessment (ASA).

The frameworks evaluated here were selected from the frameworks examined in Maemura et al. (2017), based on the criteria described in Table 4. These criteria were developed to make the study relevant and valuable for the participants and the researchers.

Table 4. Selection criteria for the frameworks studied

Criterion	Description
Mode	Focus on self-assessment independent of third-party support, not external certification.
Scope	The model's scope should coincide with concerns of digital preservation in the participating organizations.
Focus	Frameworks should be oriented towards improvement through concepts of capability or maturity, rather than compliance certification.
Feasibility	Frameworks should be sufficiently compact and well documented to make it possible to complete an assessment and study the results.

We selected the Digital Preservation Capability Maturity Model (DPCMM) and the Assessing Institutional Digital Assets (AIDA) toolkit. In the AIDA model, 31 dimensions are grouped into three areas: Technology, Organization and Resources. For each dimension, the organization as a whole and the specific department are placed into one of five stages. Illustrative statements are used to help the assessors identify the appropriate level for their department and for the whole organization. In contrast to AIDA, the DPCMM draws heavily from the OAIS Reference Model and ISO 16363 (ISO, 2012), both central frameworks in digital preservation. Figure 4 shows its overall structure, which rates the organization on five levels for each of 15 components (Dollar & Ashley, 2015). For the researchers, its claim

to be a CMM applicable across digital curation domains (archives, libraries and corporations) made it especially relevant. For the case study participants, it was particularly interesting as a low-effort model that claims to provide a comprehensive assessment aligned with the OAIS.

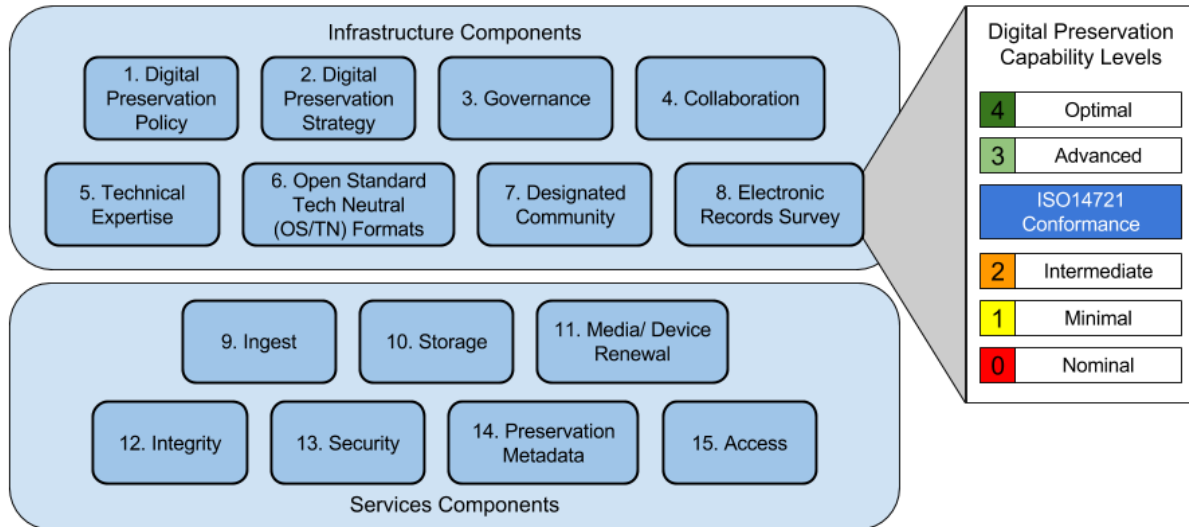


Figure 4 DPCMM components (redrawn from Fig.1 (p. 9) and Fig. 2 (p. 12) in Dollar & Ashley, 2015)

In the analytic evaluation, two analyses were performed. First, each researcher performed an independent analysis of each model using the full set of design principles as a checklist (Pöppelbuss & Röglinger, 2011). We documented the rationale for each element of the evaluation, backed by direct references to the model's documentation. Individual findings were discussed until consensus was reached between all three researchers. Second, a similar procedure was followed for the design process using the process requirements (Table 1) as a checklist, with the cautious perspective that limited inferences can be drawn about procedures from their documentation.

In the empirical evaluation, each organization used each framework in an assessment project that evaluated the organization's digital curation capabilities. Each organization therefore completed two assessment projects. The researchers studied each project using the data collection methods described in Figure 2. Figure 5 summarizes the case study design. In the comparative analysis of frameworks, the 2x2 setup supported the comparison of frameworks across organizational contexts. This enabled us to distinguish factors intrinsic to the framework highlighted across organizations from contingent questions of organizational fit and unique circumstances.

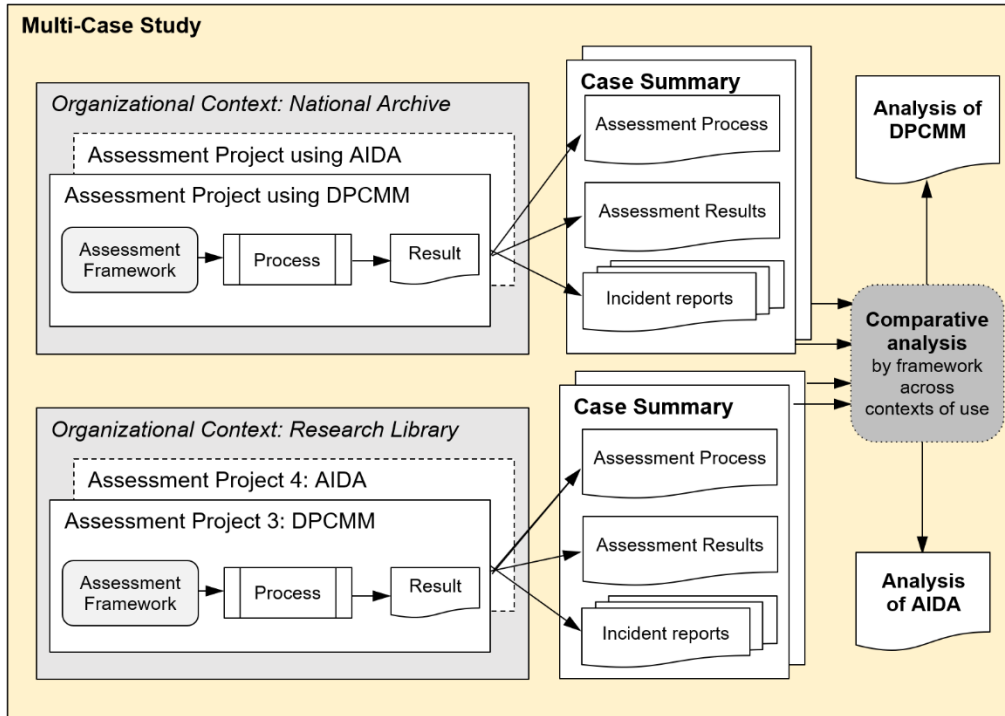


Figure 5 The Multi-Case Study Design used for empirical evaluation in our instantiation

Evaluation Results

Summary of Assessment Projects

All assessments were completed within a working day of effort. Table 5 summarizes the assessment process, incidents and results for the DPCMM projects, and Table 6 mirrors this for the projects using AIDA.

Table 5 Summary of the assessment projects using DPCMM: Process and Results

Aspect	DPCMM at ASA	DPCMM at UTL
Time spent	The assessor took an hour to review the DPCMM documentation before the assessment. The assessment was completed by the assessor, with the aid of the note-taker, in two hours.	The assessment was completed in three sessions on two consecutive days in about 4 hours.
Effort	5 person-hours	7.5 person-hours
Resources used	Most of the time was spent reading statements and explanations, DPCMM background documents and material in the accompanying tool DigitalOK.	The assessor used the documentation, User Guide and online tooltips as well as documentation from other sources.
Skills and knowledge used	The assessor rarely struggled to recall information about the organization. He frequently described a specific document or scenario to support his judgments.	The assessor was able to answer all questions with his own knowledge of the organization.
Value and use of results	The assessor found the process easier than the detailed Repository Audit and Certification checklist (ISO, 2012) and saw the results as	The assessor saw the model as useful prompt for reflection, but consistently placed more value in the process of undertaking the assessment than

	“sufficient and helpful” for an organization planning a digital repository. He found the assessment reflected reasonable strengths and weaknesses and was useful for internal advocacy for resources, and he considered the possibility of a later reassessment, but missed a clear explanation of what the levels mean and how to progress.	the results, in which he had little confidence. While he found general value in the model’s scope, he found the statements within each component too prescriptive and would prefer a model that emphasized the rationale behind each component.
Accuracy and Reliability of results	The assessor found most scores to accurately reflect his expectations, but rejected the low score in <i>Archival Storage</i> as inaccurate and justified this by describing the highly advanced mechanisms in place at the organization. He attributed the incongruence to a lack of clarity in both the model and the guidance on tool use, and to ambiguity over whether the criteria were meant to be cumulative.	The assessor found the overall results were biased too high and considered the framework incapable of capturing the range and breadth of varying capabilities that exist simultaneously within UTL. The scores were not low enough to motivate organizational changes, nor granular enough to point to specific directions to take in order to initiate change. The overall results matched his expectations, and no new insights or unexpected outcomes emerged: “...it didn’t make me think about things any differently.”

Table 6 Summary of the assessment projects using AIDA: Process and Results

Aspect	AIDA at ASA	AIDA at UTL
Time spent	The assessment was conducted by the assessor, with the aid of a note-taker, in about four hours.	The assessment was conducted by the assessor, with the aid of a note-taker, over two days. It took under three hours, plus short breaks due to unrelated technical difficulties.
Effort	7 person-hours	7 person-hours
Resources used	The assessor used a paper copy of the elements for reference, and the note-taker recorded the final stage (1-5) as it was determined for department and institution for each element, as well as rationale. Most of the assessor’s time was spent reading the criteria and determining how to apply or compare each to the situation at ASA.	The assessor used a paper copy of the elements for reference and the note-taker used the electronic templates to record the stages (1-5) for department and institution for each element. The rationale for scores were noted and confirmed by the assessor before moving on to the next element.
Skills and knowledge used	The assessor was able to recall most information required for most elements from memory. For financial auditing and policy components, he had limited knowledge of detail and noted that consulting with other staff would be necessary to provide more detailed and nuanced responses.	The assessor recalled all relevant information required to complete the assessment. More details and examples were provided for the department level, since this is the area of focus for the assessment and the practitioner.
Value and use of results	The assessor found the results generally useful for their coverage of a range of areas addressed by the model and the depth with which it examined them. AIDA served to call attention to the need for further development related to criteria in the Organization leg, which matched the assessor’s intuition about ASA’s strengths and weaknesses. However, he felt insufficient	The assessment was seen as valuable, although the results did not identify discrete targets or objectives for improvement, only broad indications of the strengths and weaknesses of different areas. The assessor found that AIDA “struck a good balance between being granular enough... without making it overwhelming”. He emphasized the model’s flexibility and exemplar

	guidance was provided on the interpretation of the results, which limited the value of the report.	statements. He liked that the framework emphasized the expertise and responsibility of the assessor in knowing the organization and encouraged informed interpretation of criteria.
Accuracy and Reliability of results	The assessor expressed modest confidence in the results, based on the relative agreement between the results and his sense of ASA's strengths and weaknesses. He expressed a willingness to share the results with colleagues and supervisors within the organization, provided they were presented with sufficient context. However, he would not share the results publicly. He noted that for some elements in Resources, additional input from other staff would help improve accuracy.	The results of the framework were generally in line with the assessor's expectations and seen as an accurate representation of the organization. He noted, "this feels pretty solid to me actually... this rings true to me." The relative scores across and within the legs of technology, resources and organization seemed appropriate to the assessor. He had modest confidence in the results, however, recognizing the potential for inaccuracy in his own scoring.

The implementation of the empirical evaluation method facilitated the structured description of each assessment project and its outcomes. Additionally, it provided twenty-two contextualized incident descriptions. Table 7 summarizes two example incidents that occurred in the use of the DPCMM.

Table 7 Example incidents in the use of DPCMM.

Framework location and description	Description of difficulty
<i>Electronic Records Survey (8)</i> addresses ways to investigate, document and manage an organization's electronic records, both for existing and projected "volume and scope of electronic records that will come into its custody." The concept of a survey is framed around the existence of retention schedules, a document specific to archives.	The assessor at UTL had to interpret electronic records concepts in a library context. He initially attempted to address the statements as-is but found it necessary to re-interpret the statements about "retention schedules" since UTL focuses on collections. He decided to interpret the statements through a mental "crosswalk" to "collections policies," while maintaining what he saw as the question's intention. The component is focused on "Electronic Records," but the assessor extended this to other digital materials being preserved. While making this judgement, he addressed the differences between retention schedules and collections policies and acknowledged that this is not an exact or perfect fit since each document would address risk differently.
<i>Access (15)</i> addresses access to Dissemination Information Packages (DIPs) by communities of "Users." DIP is a central concept of the information model of the OAIS and a mandatory element of OAIS compliance. In the OAIS, DIPs are provided to 'Consumers'.	The ASA does not currently provide direct access to external users, so the assessor at ASA interpreted the statements using archivists as the repository system's users. He interpreted statements as achieved since the system is capable of providing appropriate information packages to users, even though it is not the organization's current practice to provide these. He interpreted "Users" here differently from his interpretation in the component <i>Designated Community (7)</i> , where he considered it synonymous with the OAIS term "Consumers."

Comparative analysis of assessment projects

We illustrate the comparative analysis step of our comprehensive evaluation methodology by discussing the evaluation results for the DPCMM in detail. We provide a condensed summary of its analytic evaluation before discussing the results of the comparative analysis of DPCMM across the two contexts of use. The subsequent section summarizes the evaluation results of AIDA. The evaluation report (Becker et al., 2018) supplements this summary. It discusses the empirical evaluation of both frameworks in additional depth, presents detailed characteristics from the analytic

evaluation of each framework for each design principle and design process requirement, and documents the application of the methodology in detail to facilitate its replication.

Evaluating the DPCMM

The analytic evaluation identified multiple weaknesses in the DPCMM's design. For example, the descriptive use principle *intersubjectively verifiable criteria* requires consistency of the structure and language used in individual criteria. The criteria of DPCMM are phrased in specific terms, but inconsistencies in the terminology suggest that they may be judged differently by different assessors. In addition, many criteria are not atomic but consist of multiple individual assertions joined by "and". This composite structure might make it difficult to apply them to precisely describe the possible states of an organization's activities. The analysis further revealed several mismatches with the structure and concepts of the SEI CMM and showed that the DPCMM misappropriates its concepts. DPCMM is also not fully consistent with central concepts and constructs in its application domain. It further does not specify the nature of the entities under investigation nor define exactly how the various aspects of maturity develop over time and contribute to an organization's ability to preserve and manage digital assets. This absence of clarity about the underpinning theoretical foundations does not surface in a casual reading of DPCMM. It raises concerns about the level of robustness that can be expected from its application in diverse environments.

The in-depth analytic evaluation was iteratively brought into the empirical analysis to structure, categorize and explain the incidents of difficulties observed in the assessment projects. The analysis of these incidents highlighted characteristics of the framework around the following four themes.

Interpreting and applying criteria during an assessment was often found challenging, both in terms of domain terminology and in applying the model's structure to concrete organizational settings in the absence of helpful guidance. For example, several components contain quantifying statements including "all" conditions such as "The strategy calls for the transformation of all electronic records in native file formats to preferred preservation formats at ingest." (Digital Preservation Strategy (2)). These were interpreted very differently across organizations.

At UTL, the organization's heterogeneity meant the assessor would not confirm statements addressing "all" conditions. He noted that the types of relationships with record producers and the diversity of records producers they work with make it impossible and, in fact, unreasonable, to have complete control over or knowledge of their various record creation processes. Similarly, for questions around Designated Community and Access, UTL has structures in place to understand user groups, but could not cater to 100% of the needs of 100% of users as demanded by the top-level statements found in these components.

The assessor at ASA, on the other hand, often chose to assess the organization positively for a criterion if the "system is capable of doing it" or it's "theoretically possible" even when no active practices were currently in place or the statement was not relevant to the organization.

Inconsistencies in the logical structure of the model surfaced in both assessments and often reflected the inconsistencies identified in the analytic evaluation. For example, while most DPCMM components ask assessors to select all statements that apply to their organization, two components instead ask the assessor to select the statement that best fits their situation. Only one of the input forms ensured they would indeed select exactly one.

Because of these inconsistent affordances, the assessor at UTL did not notice the 'choose one' statement for *Preservation Metadata (14)* and selected multiple statements. The tool scored UTL at the highest level chosen, a rating deemed incorrect by the assessor.

The assessor at ASA had similar problems with the structure of the component *Archival Storage (10)*, in which the form of the statements initially begins with "Only one instance..." but shifts at higher levels. It was unclear to the assessor which statements he should select.

Terminology was often found inconsistent or incomplete, and adaptation to different contexts not always effective. As a result, terminology issues played a central role in several incidents of difficulties. For example, *Archival Storage (10)* addresses concerns such as the storage tiers used, the degree of automation in establishing a chain of custody, and the creation and use of operational statistics. For the latter, one requirement states that "Semi-automated capture

of archival storage operational statistics are used to support systematic digital preservation planning.” The terms “near-line” and “off-line” storage tiers are not defined, and no examples are provided.

The assessor at ASA found it difficult to categorize the organization’s set-up with terms such as “near-line” and “off-line” for which no definition or examples were given. Aspects of this component seemed too prescriptive and not applicable. He struggled to choose the statements with the best fit. He noted that the system is capable of capturing operational statistics but these are not currently used for systematic planning, and left these composite statements unchecked as they were only partially met.

To the assessor at UTL, it was unclear what the storage tiers might look like or how they apply in UTL’s particular case. UTL’s practices and systems seem to cover some aspects of near-line and off-line, but also overlap. The assessor interpreted some statements as how easy or quickly retrieval or recovery of information could be completed in the case of an emergency or loss.

Conceptions of maturity were criticized by both assessors. In either organization, some achievements described as ‘optimal’ by DPCMM were insufficient from the perspective of the assessors. Both felt that there should be more room for improvement at the highest level of maturity.

At ASA, the assessor specifically criticized the lack of attention to learning in the model’s concept of maturity: ‘The highest maturity level . . . should be something that is working on further optimization’, a concept central to the original CMM but not addressed in the DPCMM. He proposed incorporating concepts of optimization or continuous review to the Stage 5 requirements.

At UTL, the highest level statements were not considered adequate for the goals the organization aims to achieve. The assessor noted that there should be more room at the top: “there are certain pieces of this - digital preservation strategy being one that jumps out at me - I feel like we are not where I want to be, so I guess I’m a little surprised . . . that could be rated so highly.”

Evaluating AIDA

The evaluation results for the second framework, AIDA, showed a similar balance of resource investment, reliability, value, and use. The same four themes emerged in the evaluation of AIDA’s use across assessment projects. The characteristics of the framework surfaced in those assessments, however, highlighted the difference in AIDA’s design.

Most importantly, AIDA uses a highly illustrative structure, employs a flexible approach to scoring and provides a different scale of maturity. When issues of terminology, logical structure and structural inconsistency made it difficult to interpret criteria, AIDA’s flexible structure and guidance allowed the assessors to re-interpret the criteria as they saw fit within their context. The use of illustrative statements encouraged them to do so and provided better support than the rigid structure and scoring mechanism of the DPCMM. As a result, they talked positively about AIDA’s value, but the absence of clear thresholds and a glossary limited the intersubjective reliability of assessments.

Similarly, the maturity concept of AIDA was criticized, but for very different reasons than DPCMM’s: AIDA is based on a model that emphasizes external collaboration beyond institutional boundaries at the highest maturity stage. This was considered inappropriate. At ASA, externalization did not fit well with the institutional mandate; at UTL, it was not considered an appropriate ideal of organizational maturity.

Both assessors appreciated that AIDA prioritizes the facilitation of communication between different groups over a reproducible outcome. This suggests that for many stewardship organizations, assessment can be at least as valuable as an instrument for learning and organizational development than as a certification tool.

Research Findings

What methods can be used to evaluate assessment frameworks in digital curation?

The results above demonstrate that our evaluation methodology closes the identified evaluation gap, and the discussion showed the importance of combining analytic and empirical evaluation methods to achieve this. The empirical evaluation method ensured that multiple perspectives and sources of evidence were considered by triangulating: evidence about the organization created through the application of the assessment procedure (Figure 1); the assessors' experience as reflective practitioner, through questionnaires and interviews (Figure 2); and independent evidence of the assessment project, through observation, video recordings and a detailed analysis of the sequence of recorded events (Figure 2). However, only the further combination of empirical with analytic perspectives enabled a reflective understanding of the role of the framework in each assessment project.

The evaluation outcomes provide value to practitioners, framework designers and researchers.

For practitioners, the independent evaluation of frameworks provides robust insights into their use in specific organizational contexts that guide organizational stakeholders in choosing between frameworks. In combination with the analytic evaluation, the empirical study of frameworks in real contexts of use illuminates their characteristics and how they influence the assessment process and its results.

For framework designers, the insights into each framework's design inform potential revisions of the framework by identifying weaknesses in framework design. The empirical illustration of strengths and weaknesses in observed situations also supports future trade-off decisions in design. For example, the evaluated frameworks take opposite approaches to support descriptive use: One focuses on strictly defined sets of criteria, the other uses illustrative statements. Neither provides consistent or rigorous terminology. The study showed that different difficulties arise in each case. The flexible statements in AIDA were applied more effectively than the rigid concepts in the DPCMM. Even diagnostic self-assessments, however, require intersubjectively verifiable criteria to support the consistent characterization of organizations over time. This remains a challenge to consider.

For researchers, the methodology's rigor and transparent documentation support replication studies that accumulate insights over organizations, frameworks and time. The insights from the implementation of the methodology in turn enable us to examine the relationships between design process, product and use in the discussions of RQ2 and RQ3 below.

How does the design of an assessment framework influence its use and value in different contexts?

To examine the influence of framework design on quality in use, we analyzed the events observed in the use of frameworks in assessment projects. We used the explanatory schema introduced in Figure 3 as a template to relate product quality to quality in use by linking an incident to a product characteristic it highlights and a set of principles this relates to, and to a consequence and its effect on perceived quality.

The results use design principles to explain characteristics of the frameworks observed in practice. For example, the pattern shown in Figure 6 describes the two incidents observed in the DPCMM projects listed in Table 7. Both directly impacted the efficiency, reliability and effectiveness of assessments. In this recurring pattern, a specific term is left undefined in the model. During an assessment project, assessors encounter difficulties interpreting the term and search for its definition. Without a definition, they interpret it based on their own background experience. In principle, this characteristic could have been uncovered during analytic evaluation by a thorough reading of the documentation guided by the structured set of principles. However, such details often go unnoticed during development. Only observing the practice of assessment allows us to identify particularly relevant product characteristics and understand their consequences in real environments. This empirical analysis of the *product in use* then serves to focus and strengthen the analysis of *the product itself* and support improvements in its design.

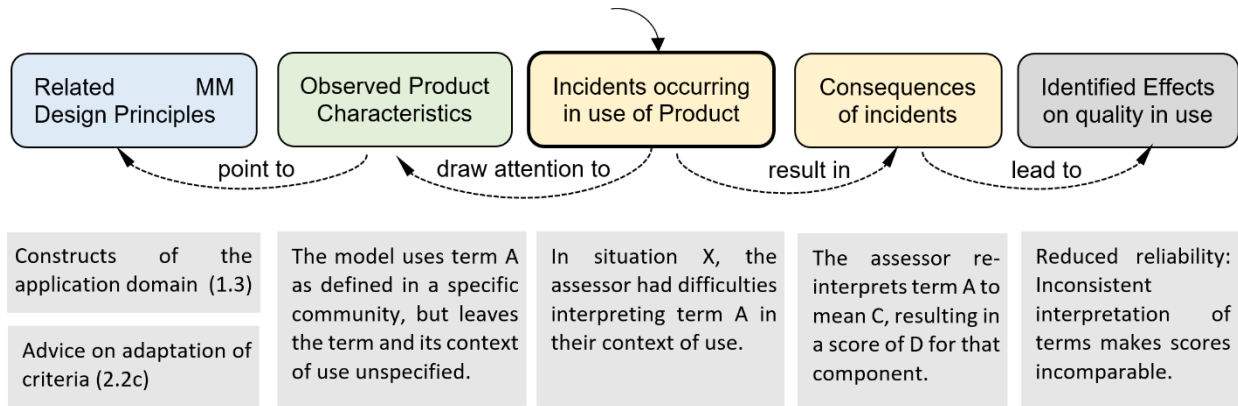


Figure 6 Relationships of design principles, quality characteristics, events observed in incidents, and effects

The product characteristics highlighted in this pattern show that design principles about terminology and adaptation identify aspects of the design that should be improved. Table 8 summarizes highlighted characteristics and the principles they relate to. The evaluation report discusses five such incidents.

Table 8. Interpreting criteria and terminology is found challenging for the DPCMM

Characteristics of the model	Relevant design principles
DPCMM claims to be flexible, but no advice is provided on how to adapt or re-interpret criteria.	application domain (1.1a)
Terms such as "electronic records" and "retention schedules" are used in DPCMM in a document targeted (among others) at libraries and companies. The framework provides no guidance for translating the terms to these domains.	central constructs of the application domain (1.3) Advice on adaptation and configuration of criteria (2.2c)

How can the evaluation of assessment frameworks inform future design methods?

We draw several conclusions that inform the design of future assessment frameworks.

First, our findings show that both analytic and empirical methods are needed to perform the evaluation that must be part of the design process when successive iterations of models are built (Becker et al., 2009; Peffers et al., 2007).

Second, the links between framework characteristics and design principles established above highlight how the characteristics of a conceptual framework designed to be used in an organizational assessment influence its quality in use. The findings thus establish the central relevance of maturity model design principles for assessment frameworks in digital curation. For example, a framework must define the central constructs of the application domain and document them comprehensively in a robust vocabulary. This is even more important if the framework is to be applied consistently across different communities: While many concerns in digital curation converge across research data management units, archives, libraries and museums, the terminology applied in practice does not. Similarly, it is essential to develop a robust understanding of what exactly is being assessed and to distinguish technical capability, process maturity, responsibility, accountability, and other aspects of organizational maturity. To make assessment results valuable in practice, a baseline diagnosis should support prescriptive use to identify possible improvement measures. Later reassessments then contribute to building a knowledge base of improvement efforts and their effects. The evaluated frameworks do not address these needs, and the study shows that this limits their value in practice.

Finally, unpacking the events discussed above to determine characteristics of frameworks revealed that issues generally occurred because of several factors. This demonstrates that design principles should not be applied in isolation. For example, a revision of the DPCMM could certainly address the individual gaps of terminology and structural inconsistency that we highlighted. But overall, an attempt to address levels of maturity and maturation paths and define clear evaluation criteria for descriptive use can only be successful if it can also build on clarity about the nature of the entities under investigation and a robust understanding of how these entities change in the application context. Since these concepts provide the foundations of a model's design, design principles must be considered carefully throughout the design of assessment frameworks. The study thus allows us to distill the lessons learned in the evaluation into concrete guidelines for future research and design practice, structured by the requirements described in Table 1.

- R1. **Comparison with existing maturity models:** To provide effective support to organizations, the design of future frameworks should focus on supporting improvement over time based on repeated assessments.
- R2. **Iterative Procedure:** Design iterations and rationales should be fully documented to make explicit the lessons learned from analytic evaluation and empirical study of the model's use. Case studies of assessment projects should complement analytic evaluation by the design team.
- R3. **Evaluation** should combine the perspectives of process, product and use. Beyond expert review, empirical evaluation is essential. It studies the assessment process itself and identifies incidents that illustrate deficiencies in quality in use that manifest in subtle ways. Evaluation should avoid confirmation bias; support iterative improvements of frameworks; and provide transparent evaluation reports to support independent scrutiny.
- R4. **Multi-methodological Procedure:** The analytic evaluation of a framework can uncover important issues and identify conceptual weaknesses. The empirical evaluation in real-world usage environments similarly can unearth quality issues. The two perspectives are mutually complementary and must be applied in combination. Design methods should incorporate key principles throughout the design process.
- R5. **Identification of Problem Relevance:** Frameworks should be designed with full consideration of the purposes of application. Support for focused improvement efforts through repeated self-assessment over time is currently missing. A clear assessment procedure and guidance is needed.
- R6. **Problem Definition:** Assessment frameworks should define their intended purpose and context of assessment, the concepts of capability and maturity, and their underpinning foundations.
- R7. **Presentation of Results:** Frameworks should specify explicitly the objectives of assessment and provide realistic guidance on a model's scope and anticipated context of use.
- R8. **Scientific documentation:** Precise documentation supports the evaluation and iterative design of assessment frameworks. The methodology presented in this article supports rigorous evaluation and documentation and provides guidance for future research.

Conclusions

This article addresses the lack of evaluation methods for assessment frameworks in digital curation. The work presented here synthesizes existing principles and methods for maturity model design into a new methodology that supports the systematic development of assessment frameworks in digital curation by evaluating their design and use. It provides a replicable structured empirical method to study the process of assessment using a given framework to reveal how it influences the assessment process. It combines this with an analytic evaluation of the design process through reviewing documentation, and of the framework itself, by analyzing its model and documentation.

The results demonstrate the value of combining empirical and analytic methods. Their combination enabled us to analyze the conceptual model presented in each framework, applied to an organization, as implemented and supported in the tool and articulated in the documentation, observed through the process of assessment and reviewed through the perspective of the assessor. The systematic evaluation of frameworks provides insights to the designers of the evaluated frameworks that they can consider in future revisions, and practical insights and words of caution to

organizations keen on diagnosing their abilities. Comparing the findings across frameworks supported the development of guidelines for the design process and resulted in methodical guidance for researchers in the field.

It is now time to apply this evaluation methodology to other frameworks and organizational contexts; integrate the evaluation with iterative design methods; and eventually, support the prescriptive use of assessment frameworks through concrete methods and conceptual models that support decision makers in identifying and prioritizing improvement measures to advance their organization's abilities.

References

- Becker, C., Maemura, E., Moles, N., Whyte, J., Kulovits, H. & Marks, S. (2018). *Evaluating Assessment Frameworks in Digital Curation*. University of Toronto. Retrieved from <http://hdl.handle.net/1807/93425>
- Becker, J., Knackstedt, R., & Pöppelbuss, D.-W. I. J. (2009). Developing maturity models for IT management. *Business & Information Systems Engineering*, 1(3), 213–222.
- Dobratz, S., Rödiger, P., Borghoff, U. M., Rätzke, B., & Schoger, A. (2010). The use of quality management standards in trustworthy digital archives. *International Journal of Digital Curation*, 5(1), 46–63.
- Dollar, C., & Ashley, L. (2015). *Digital Preservation Capability Maturity Model (DPCMM) Background and Performance Metrics Version 2.7*. Retrieved from http://www.securelyrooted.com/s/DPCMM-Background-and-Performance-Metrics-v27_July-2015-26m4.pdf.
- Eisenhardt, K. M., & Graebner, M. E. (2007). Theory building from cases: opportunities and challenges. *Academy of Management Journal*, 50(1), 25–32.
- Frick, N., Küttner, T. F., & Schubert, P. (2013). Assessment methodology for a maturity model for interorganizational systems – the search for an assessment procedure. In *2013 46th Hawaii International Conference on System Sciences (HICSS)*. IEEE.
- Garvin, D. A. (1984). What does “product quality” really mean? *Sloan Management Review*, 26(1), 25-43.
- Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, 37(2), 337–356.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.
- ISO. (2012). *ISO 16363 - Space data and information transfer systems - Audit and certification of trustworthy digital repositories*. International Standards Organization.
- ISO. (2001). *ISO/IEC 9126-1:2001 - Software engineering - Product quality - Part 1: Quality model*. International Standards Organization.
- ISO. (2011). *IEC 25010: 2011: Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models*. International Standards Organization.
- Jokela, T., Siponen, M., Hirasawa, N., & Earthy, J. (2006). A survey of usability capability maturity models: implications for practice and research. *Behaviour & Information Technology*, 25(3), 263–282.
- Maemura, E., Moles, N., & Becker, C. (2017). Organizational assessment frameworks for digital preservation: a literature review and mapping. *JASIST*.

- Maier, A. M. (2012). Assessing organizational capabilities: reviewing and guiding the development of maturity grids. *IEEE Transactions on Engineering Management*, 59(1), 138–159.
- March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, 15(4), 251–266.
- McHugh, A., Ross, S., Innocenti, P., Ruusalepp, R., & Hofman, H. (2008). Bringing self-assessment home: repository profiling and key lines of enquiry within DRAMBORA. *International Journal of Digital Curation*, 3(2), 130–142.
- Mettler T., Rohner P., Winter R. (2010). Towards a classification of maturity models in information systems. In: D'Atri A., De Marco M., Braccini A., Cabiddu F. (eds) *Management of the Interconnected World*, 333-340. Physica. https://doi.org/10.1007/978-3-7908-2404-9_39
- Miller, L., Blake, M., & Sorsby, M. (2012). Evaluation of the assessing institutional digital assets (AIDA) toolkit. *Science & Technology Libraries*, 31(1), 92–99.
- Paulk, M. C., Curtis, B., Chrissis, M. B., & Weber, C. V. (1993). *Capability Maturity Model for Software, Version 1.1* (CMU/SEI-93- TR-24). Pittsburgh, Pa.: Software Engineering Institute.
- Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3).
- Pöppelbuss, J., & Röglinger, M. (2011). What makes a useful maturity model? A framework of general design principles for maturity models and its demonstration in business process management. *19th European Conference on Information Systems (ECIS)*, Helsinki.
- Röglinger, M., Pöppelbuß, J., & Becker, J. (2012). Maturity models in business process management. *Business Process Management Journal*, 18(2), 328–346.
- SCAMPI Upgrade Team. (2011). *Standard CMMI Appraisal Method for Process Improvement (SCAMPI) A, Version 1.3: Method Definition Document*.
- Sinclair, P., Duckworth, J., Jardine, L., Keen, A., Sharpe, R., Billenness, C., Humphreys, J. (2011). Are you ready? Assessing whether organizations are prepared for digital preservation. *International Journal of Digital Curation*, 6(1), 268–281.
- Smallwood, N., & Ulrich, D. (2004). Capitalizing on capabilities. *Harvard Business Review*, 82(6), 119-127.
- Wendler, R. (2012). The maturity of maturity model research: a systematic mapping study. *Information and Software Technology*, 54(12), 1317–1339.
- Yin, R. K. (2013). *Case study research: design and methods*. Sage Publications.