

Assessing Digital Preservation Capability Using a Maturity Model Process Improvement Approach

Charles M. Dollar and Lori J. Ashley

February 2013

Abstract

The public and private sector have recognized over the last decade that the systematic management of their digital assets requires implementing a program that ensures on-going access to authentic, usable digital records that have long-term¹ operational, regulatory, legal, or cultural memory value. The Open Archival Information System (OAIS) Reference Model (ISO 14721) identifies high level services and requirements that a trustworthy repository should provide to support long-term access. Additional standards (i.e., ISO 16363) specify auditing criteria for the certification of trustworthy repositories. Both standards are notable contributions to the emerging field of digital preservation but they have several implementation limitations.

In this paper we introduce a Capability Maturity Model (CMM) that organizes the digital preservation requirements of the ISO Standards into fifteen components with metrics to assess maturity. The model is a tool for charting the evolution from disorganized and undisciplined management of digital records, or the lack of a systematic digital continuity approach, into increasingly mature stages of digital preservation capability.

The goal of our Digital Preservation Capability Maturity Model (DPCMM) is twofold:

1. To help practitioners identify at a high level the capabilities of their organization relative to optimal digital preservation capabilities; report gaps, capability levels, and preservation performance metrics to resource allocators and other stakeholders; establish priorities for achieving enhanced capabilities to preserve and ensure access to long-term digital records; and
2. To focus attention on digital continuity as a discipline for proactively addressing digital preservation issues at or near the time of the capture or creation of digital records of long-term value.

What Is A Capability Maturity Model?

The Software Engineering Institute of Carnegie Mellon University released the Capability Maturity Model for Software (CMM or SW-CMM) in 1990. The CMM was developed to enable organizations to assess the maturity of their software development processes and identify key practices necessary to improve the capability of those processes.² The CMM defines five progressive stages of process maturity³ based on an organization's support for certain key software development areas that are described generally as Initial, Repeatable, Defined, Managed, and Optimized. Each stage includes a series of associated activities and baseline metrics used to measure performance. These maturity stages are cumulative: an organization achieving a higher stage of maturity must implement and sustain all of the requirements for that stage in addition to requirements for all of the lower stages. This capability maturity model has been adapted for human resources,⁴ system engineering,⁵ software acquisition,⁶ technology investment,⁷ enterprise architecture,⁸ and records management,⁹ among others.¹⁰

Why A Digital Preservation Capability Maturity Model?

The Open Archival Information System (OAIS) Reference Model (ISO 14721) identifies high level services and requirements that an archive should provide to support long-term access. An additional standard (i.e., ISO 16363) specifies auditing criteria for the certification of trustworthy repositories. Both standards are notable contributions to the emerging field of digital preservation but they have several implementation limitations. The objective of ISO 14721 is to serve as a reference model. The OAIS defines digital preservation services and associated activities at a very high level. OAIS services and activities must be deconstructed into terms that are readily understood and can be applied in operational archival environments. In contrast, the audit criteria certification checklist of ISO 16363 includes more than one hundred requirements, and conducting an audit presumes an external audit team is on site and authorized to certify the repository. An audit typically involves months of preparation to acquire and organize documentation. More importantly, neither standard identifies explicit performance metrics to assess the current digital preservation capabilities of repositories (digital archives) or information systems that may act as surrogate trustworthy repositories. Both standards imply a "one size fits all" approach to ensuring long-term access to authentic digital records. Finally, neither standard explicitly supports an incremental digital preservation improvement capability plan.

The Digital Preservation Capability Maturity Model (DPCMM)¹¹ presented in this paper draws upon the overall framework of the CMM development model but is not intended to be a rigorous model with precisely defined parameters. The DPCMM is a systematic tool to chart the evolution from a disorganized and undisciplined electronic records management program, or one that does not exist, into increasingly mature stages of digital preservation capability. The DPCMM is designed to help identify, protect and provide access to long-term and permanent digital assets. The goal of the DPCMM is to support the management of a digital preservation program that:

- Identifies and monitors at a high level where the program is in relation to an optimal digital preservation program.
- Establishes priorities and an improvement roadmap to achieve enhanced digital preservation capabilities over time.
- Reports digital preservation capability gaps and achievements to resource allocators and stakeholders.

Stages of Digital Preservation Capability Maturity

The Digital Preservation Capability Maturity Model displayed in Figure 1 has five stages that track closely with the five stages of the CCM discussed earlier, albeit with a specific digital preservation emphasis.

Stage 1: (Nominal) a systematic digital preservation program has not been undertaken and **most, if not all, electronic records that merit long-term retention are at risk.**

Stage 2: (Minimal) digital preservation capabilities are rudimentary and do not rise to the level of ISO 14721/ISO 16363 specifications. **Consequently, most electronic records that merit long-term retention are at risk.**

Stage 3: (Intermediate) the organization supports ad hoc initiatives and projects that approach but do not conform fully to ISO 14721/ISO 16363 specifications. There is an established basis for proactive and sustainable digital preservation improvement actions over time. **Nevertheless, it is likely that some electronic records that merit long term retention remain at risk.**

Stage 4: (Advanced) the organization has a robust infrastructure and the preservation of electronic records is undertaken with a framework that conforms to most of the ISO 14721 specifications and the criteria of ISO 16363. **Few electronic records that merit long-term preservation are at risk.**

Stage 5: (Optimal) represents the highest level of sustainable conforming ISO 14721/ISO 16363 digital preservation capability and repository” trustworthiness” that an organization can achieve. **No records that merit long-term retention are at risk.**

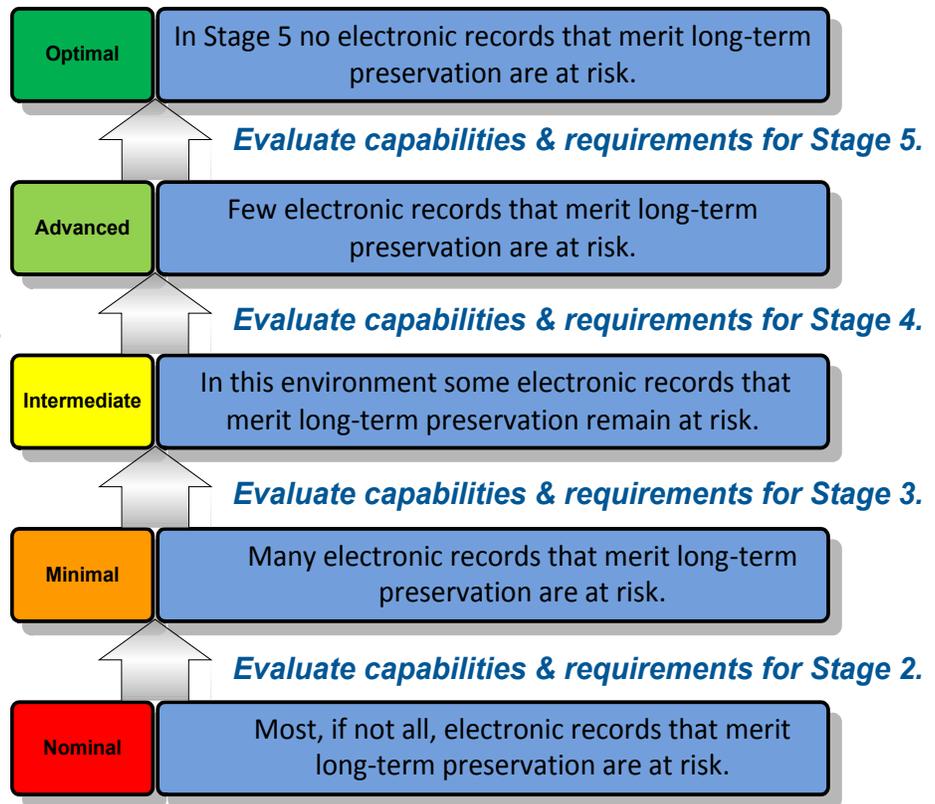


Figure 1. Stages of Digital Preservation Capability Maturity

Digital Preservation Capability Maturity Model Components

The Digital Preservation Capability Maturity Model (DPCMM) consists of three interdependent digital preservation domains: infrastructure, one or more digital repositories, and services. Figure 2 displays the fifteen components of the DPCMM and their relationship to the top level domains.

These components of the DPCMM are an amalgamation of key specifications, requirements, and activities abstracted from ISO 14721 and 16363 standards and digital preservation “best practices”.¹² Two major stakeholder groups – records producers (or donors) and users who seek access to the contents of the digital repository, also appear in the model.

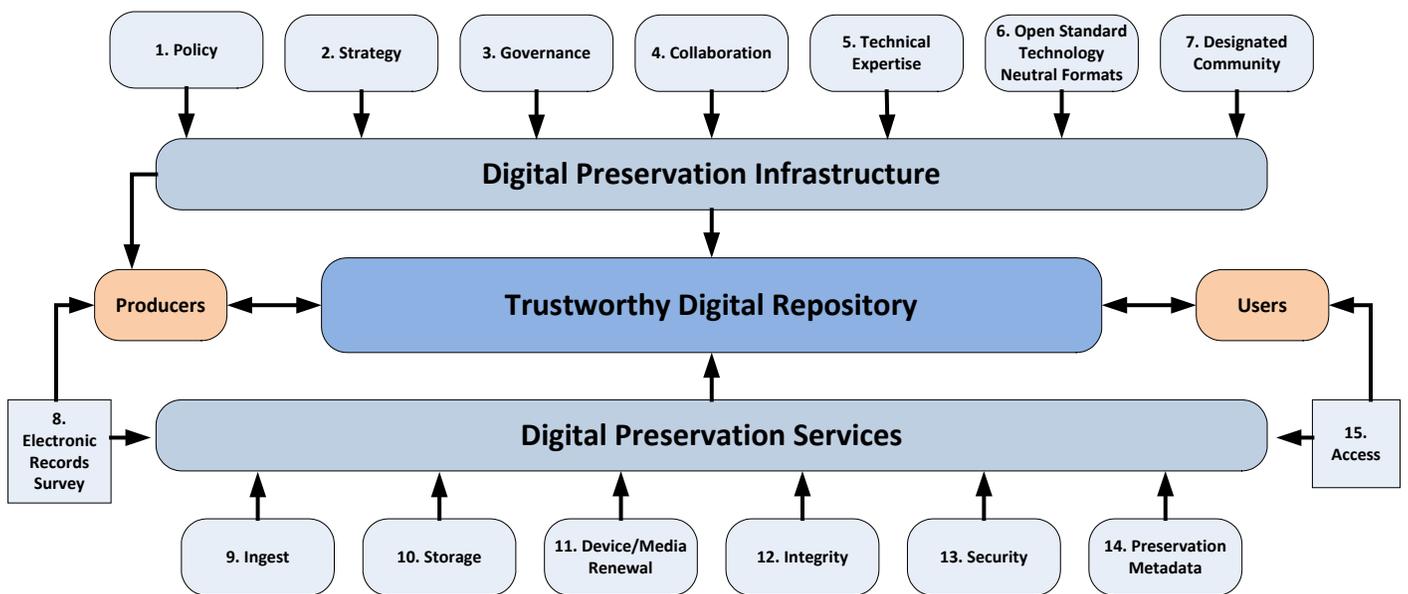


Figure 2. Digital Preservation Capability Maturity Model

Digital Preservation Infrastructure features seven (7) components that are essential to ensuring sustained organizational commitment including, human, technical and financial resources, to the long-term preservation of digital records that are created, received or acquired by the organization. The Digital Preservation Infrastructure components are:

- Digital Preservation Policy
- Digital Preservation Strategy
- Governance
- Collaboration
- Technical Expertise
- Open Standard Technology Neutral (“OS/TN”) Formats
- Designated Community

Digital Preservation Services include eight (8) components that are required for continuous monitoring of external and internal environments to plan and take necessary preservation actions that sustain the integrity, security, usability and accessibility of digital records stored in repositories. The Digital Preservation Services are:

- Electronic Records Survey
- Ingest
- Archival Storage
- Media/Device Renewal
- Integrity
- Security
- Preservation Metadata
- Access

The DPCMM describes the scope and focus of each of the fifteen components. The Electronic Records Survey component (*see below*) is a critical interface between Records Producers, the repository, and the Digital Services domain. The Electronic Records Survey addresses the need for an informed estimate of the volume, file formats, and types (e.g., images, text, and databases) of digital content that will be transferred to the digital repository or safeguarded by record producers in their own technology environments. The Electronic Records Survey component is also dedicated to mitigating technological obsolescence at the time of records transfer to the trustworthy repository. Practitioners are encouraged to engage records producers to capture preservation-ready digital records at or near the time of their creation or receipt. Below is the DPCMM description of the Electronic Records Survey component.

DPCMM COMPONENT 8: Electronic Records Survey

Each organization is responsible for records created, received or acquired that are evidence of its business activities, regardless of the format or media used, and must ensure the records' authenticity, integrity, usability and reliability for as long as they are required. Records with long-term retention requirements or archival (permanent) value are often transferred to the custody of a centralized Records Management and/or Archives function for preservation. Due to the fragility of electronic records, organizations are advised to proactively address digital preservation as close to the time of electronic records creation or capture as practicable. This can only be accomplished if the organization has a comprehensive inventory of electronic records as well as collaborative working relationships and agreements between stakeholders that include Records Producers, Legal/Compliance, Archives, Records Management, Information Technology/Services as well as third party application, solution and service providers.

A key feature of a conforming ISO 14721 open archival information system is the reliance on open standard interoperable technology neutral formats. During Ingest electronic records in proprietary formats must be transformed into formats that the organization and/or repository have adopted. Over time and with increasing volumes of electronic records, format transformation during the Ingest process may become burdensome. This obligation can be mitigated in part if "preservation-ready" records, that is, records that are in open standard interoperable technology neutral formats, are made at or near the time Records Producers create or capture the records.

The objective of an Electronic Records Survey is to identify three broad categories of records with a retention period of ten (10) years or more in order to support planning and preservation activities.

- Category 1 encompasses "**Preservation-Ready**" electronic records.
- Category 2 includes "**Near Preservation-Ready**," that is electronic records in formats for which tools are available that can export native format documents to open standard interoperable technology neutral formats.
- Category 3 encompasses "**Legacy**" records, that is, electronic records in a proprietary native format for which no export or viewer technology tools exist. Transformation of proprietary native formats into open standard, interoperable, and technology neutral formats is likely to require writing code to support this transformation, which in turn is likely to be costly.

The collection and analysis of data for an Electronic Records Survey can be accomplished by a variety of means including: web enabled surveys of record producing units and service providers, interviews with selected business units or third parties that routinely create, receive or acquire electronic records, review of the organization's records retention and disposition schedules, analysis of the organization's information technology portfolio and strategic plan, as well as the use of "network crawler" functionality to identify specific file formats currently used in the capture and storage of electronic records on network drives.

Digital Preservation Surrogates and Thresholds

ISO 14721 and ISO 16363 are the "gold standard" for digital preservation. While many organizations will strive to implement and sustain a fully conforming ISO 14721/16363 digital preservation environment, the reality is that some organizations cannot or choose not to implement a traditional standalone trustworthy repository. Reliance on routine operational environments for long-term storage is the usual alternative to a standalone repository. Other organizations lack a sufficiently mature information infrastructure and architecture, and/or have such limited technical and financial resources, that in the short run they cannot aspire to implementing a conforming ISO 14721/ISO 16363 digital archival repository.

In the meantime some organizations have accessioned "born digital" or scanned digital images through manual or semi-automated workflows. Others are addressing some digital preservation requirements with tools and services such as Contentdm®, Archivelt, BagIt, LOCKSS, and DSpace. In the public sector, grant-based projects and state-level collaboration in federally funded database and email archiving projects has been underway for years. These tools, services and projects are noteworthy despite not being in full conformance with ISO 14721 specifications. They are substantive and represent important emerging capabilities and market recognition of digital preservation requirements and standards. A handful of organizations is currently building or are testing digital repository technologies that are likely to conform to the ISO 14721/ISO 16363 specifications.

The DPCMM takes into account this spectrum of digital preservation surrogates by distinguishing between ISO 14721 conforming and partially-conforming capabilities and services by incorporating two concepts: 1) surrogate trustworthy digital repositories, and 2) digital preservation thresholds. A surrogate trustworthy digital repository as defined by the DPCMM refers to a range of services, tools, projects and initiatives currently used to address digital preservation requirements that are substantive and represent evolving/emerging capabilities.

A surrogate trustworthy digital repository may not fully or even explicitly comply with ISO 14721/16363 specifications. The DPCMM maps threshold statements to Stage 1 and Stage 2 performance metrics for each of the fifteen components to assist practitioners and other stakeholders to understand and apply these concepts.

Digital Preservation Performance Metrics

The five incremental maturity levels described previously comprise performance metrics for each component of the Digital Preservation Capability Maturity Model. The performance metrics of each Digital Preservation component constitute a checklist that is used to conduct a self-assessment of an organization's current digital preservation capability vis-a-vis that of an optimal capability. They also serve to raise awareness and educate stakeholders about current and evolving operational digital preservation practices, technology solutions, and standards.

The performance metrics for the Electronic Records Survey Component are provided below.

Level	Capability Description
0	The organization has little or no capability or resources to collect and analyze information about the volume, location, media, format types, and life cycle management requirements for electronic records.
1	The organization relies on existing retention schedules to identify electronic records of permanent historical, fiscal, and legal value in the custody of Records Producing units. It may also conduct ad hoc, one-time interviews and surveys to identify other electronic records of permanent historical, fiscal, and legal value.
2	The organization uses systematic interviews, surveys, and retrospective analysis of existing retention schedules to identify electronic records of permanent historical, fiscal, and legal value in the custody of selected records producing units. This may be enhanced by focusing on identifying “at risk” electronic records in the custody of selected Records Producing units.
3	The organization supplements analysis of “at risk” electronic records through collection of information about the volume and location (e.g., shared drives, databases, applications), media and format types of electronic records of long-term and permanent historical, fiscal and legal value in the custody of Records Producing units. The organization has identified preservation-ready and non preservation-ready electronic records in the custody of <i>most</i> records producing units.
4	The organization has identified preservation-ready and non preservation-ready permanent electronic records in the custody of <i>all</i> Records Producing units. It uses this information along with other information collected from Records Producing units to systematically manage the transfer and ingest of electronic records.

Figure 3. Electronic Records Survey Component – Performance Metrics

Digital Preservation Capability Assessment

Each of the 75 capability statements in the Digital Preservation Capability Maturity Model has an integer value ranging from 0 to 4. Using the previously described Electronic Records Survey as an example, an organization that relies solely on its record retention schedules to identify long-term and permanent records to be transferred to their digital Archives would yield a score of “1.” This score becomes the **index value** for the organization’s current Electronic Records Survey capability. Performing this assessment for all of the 15 components of the DPCMM produces an Aggregated Digital Preservation Capability Index Score that is mapped to the appropriate level of digital preservation capability.

Capability Level	Index Score
Nominal	0
Minimal	Between 1 and 15
Intermediate	Between 16 and 30
Advanced	Between 31 and 45
Optimum	Between 46 and 60

Table 1. Digital Preservation Capability Level Assessment

Both the Digital Preservation Index score for each component as well as the Aggregated Digital Preservation Capability Index Score function as a high level assessment. An organization can use the assessment results to measure its status against peer organizations as well as to develop a roadmap for incremental capability improvement. The improvement roadmap should take into account available resources and on-going initiatives and may help mitigate near-term risk exposure on some, but not all, of the components. This is an important consideration in designing an incremental digital preservation plan that is suited to the mission and designated communities of stakeholders. It is likely that constrained resources will require the prioritization of some components where significant improvement may be achieved while other components by default may undergo little improvement for the foreseeable future.

Interestingly, use of the DPCMM gap analysis checklist/performance metrics methodology thus far by the authors has raised individual and organizational awareness of the importance of digital preservation, identified interdependencies between and among various stakeholders, and sparked debate and dialogue. The assessment raises issues about the desired future state of an organization’s digital preservation capabilities and the level of risk its leadership is willing to take on. In many instances, this is likely to come down to the question of what constitutes digital preservation that is “good enough” to fulfill the organization’s mission and meet the expectations of its stakeholders within its constrained resources. This is a critical issue that the digital preservation community and those who depend on access to long-term digital records and cultural resources need to confront.

Case Study

In July 2011, the Council of State Archivists¹³ (“CoSA”) launched an initiative focused on improving efforts to manage, preserve, and provide access to U.S. state government electronic records nationwide. The goal of Phase 1 of the State Electronic Records Initiative (SERI) was to create a profile of electronic records programs in order to develop an action plan that addresses the needs of state archives and records management programs and identifies next steps.

CoSA compiled information on electronic records management and digital preservation programs as part of the SERI Phase 1 initiative. Responses to questions and transcripts from phone interviews with the directors and electronic records staff were collected from 55 state and territorial archives. CoSA invited Charles Dollar and Lori Ashley, the developers of the DPCMM, to analyze the survey results and map the findings to the fifteen (15) components of the model. In addition to providing a composite “score” on the readiness of each state and territory archives to preserve long-term and permanent electronic records, the analysis highlighted current good practices as well as enormous gaps. The consultant report stated that “Almost one-half (21) of the responding states/territories (48) registered an absolute Nominal digital preservation capability index score on each of the fifteen key process areas.”¹⁴ In November 2011, Julia Marks Young, the President of CoSA, included excerpts from this analysis at a meeting of the National Historical Publications and Records Commission.

Subsequently in 2012, the Institute of Museums and Library Services (IMLS) awarded CoSA a three-year \$500,000 grant to identify training needs and priorities for state archives, organize and conduct training programs, and to benchmark the effectiveness of the program. The program called for each state archives to take a self-assessment survey and establish a base-line digital preservation capability score. At the end of the grant program each state archives will take the self-assessment survey again and thereby document its improved digital preservation capabilities.

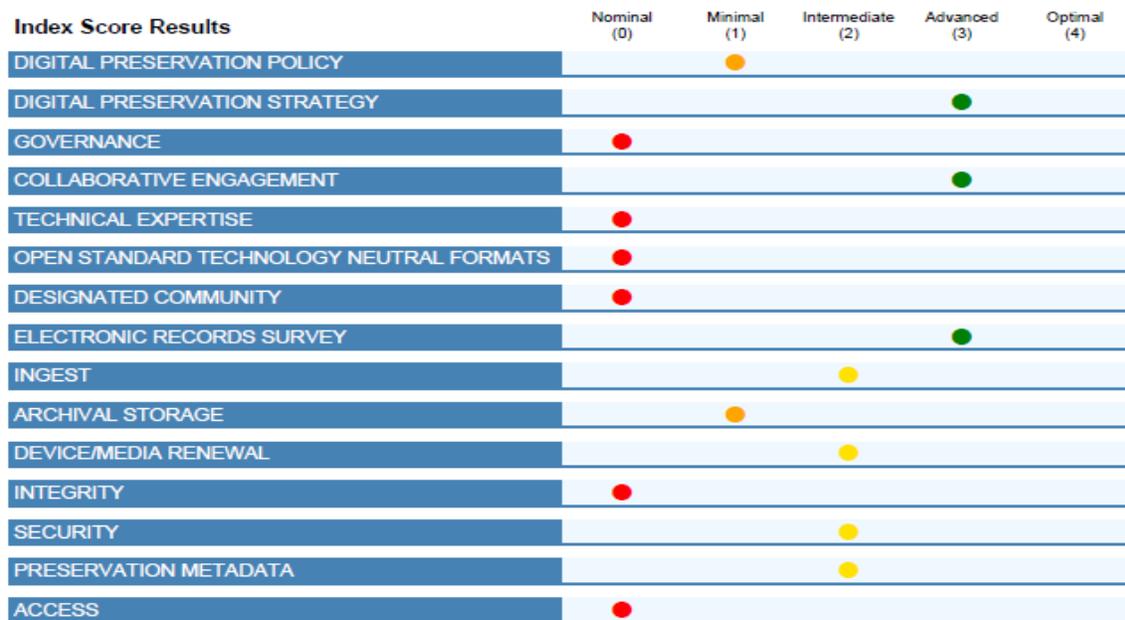


Figure 4. Aggregated Digital Preservation Index Score for a U.S. State Archives

Figure 4 is the baseline digital preservation capability scorecard of one of the state institutions that participated in the June 2012 CoSA Digital Preservation Capability self-assessment survey. The Aggregated Digital Preservation Index Score of 19 places this State Archives in the lower range of Minimal Digital Preservation Capability, which means that most of the electronic records that merit long term retention are at risk.

Conclusions and Outlook

The key strengths of the Digital Preservation Capability Maturity Model (DPCMM) - mapping of the specifications and requirements of ISO 14721 and ISO 16363 to fifteen components and the identification of incremental levels of digital preservation capability with explicit performance criteria - appear to be gaining traction within sectors of the digital preservation community. A major municipality in North America used the DPCMM to develop a five year digital preservation strategy and improvement road map. A global pharmaceutical company used DPCMM to benchmark the conformance of its digital preservation program to the specifications and requirements of ISO 14721 and ISO 16363. In June 2013 the Annual Meeting of the International Council on Archives Section of International Organizations (SIO) will include a one-day workshop based on the Digital Preservation Capability Maturity Model and assessment methodology.¹⁵ Organizers of the workshop are hoping to incorporate use of the DPCMM into the SIO's four-year educational program.

The model is a featured part of savingthedigitalworld.com, a website dedicated to fostering communication and collaboration among stakeholders who are working to address long-term digital continuity and preservation challenges. Additional background information about the DPCMM, a generic self-assessment survey, and other case studies will be published on this site in 2013.

Founding members¹⁶ of savingthedigitalworld.com hope that it will become a resource for individuals and organizations interested in understanding digital preservation issues and learning from the experience of others. In addition, the founders hope that the site will become a forum for on-going dialogue about the opportunities for and impediments to ensuring long-term access to digital content that has legal, regulatory, business, and cultural memory value.

Acknowledgements

The genesis of this Digital Preservation Capability Maturity Model Methodology is rooted in a presentation given to the Arizona Digital Records Management Task Force in 2002. Introduction to the potential use of a digital records management capability maturity model by Timothy Sprehe led to significant enhancements. The Digital Preservation Capability Maturity Model performance metrics presented in this paper were inspired in part by material developed by the International Records Management Trust to support an assessment of an organization's readiness to undertake a digital records management program.

The first use of DPCMM was in a 2007 project at the State of Delaware Public Archives. The DPCMM Methodology has undergone significant enrichment since its first formal use, the most notable being the decision of the Council of State Archivists (CoSA) to support adaptation of the model to a web-based digital preservation capability survey for fifty-six state and territorial archives. Gary Miller (Wind Lake Solutions), Richard Pearce-Moses (Clayton State University), Milovan Mistic (World Intellectual Property Organization) and Ton Bezemer (Anth.P.Bezemer LLM, The Netherlands) provided valuable commentary on the DPCMM and during development of the CoSA Digital Preservation Capability Self-Assessment.

Notes

¹ Long-term is a period of time long enough for there to be concern about the impacts of changing technologies on information held in a digital repository. This can be as short as five to seven years and extends indefinitely. In this document long-term is assumed to be 10 years or more (10+ years).

² In 2002 the Software Engineering Institute replaced CMM with a new product, Capability Maturity Model Integrated (CMMI), and discontinued support of CMM. CMMI is a robust but generic business process improvement model.

³ A good source for a description of the CMM is Mark C. Paulk, et al, "The Capability Maturity Model: A Summary," which can be accessed through the Carnegie Mello Research Showcase at <http://repository.cmu.edu>

⁴ <http://www.sei.cmu.edu/reports/01mm001.pdf>

⁵ <http://www.sei.cmu.edu/reports/96hb004.pdf>

⁶ <http://www.sei.cmu.edu/library/abstracts/reports/02tr010.cfm>

⁷ U.S. General Accounting Office, Information Technology Investment Management: A Framework for Assess and Improving Processing Maturity, GAO-04-394Gm 2004). Available at <http://www.gao.gov/new.items/d04394g.pdf>.

⁸ U.S. General Accounting Office, Information Technology: A Framework for Assessing and Improving Enterprise Architecture Management, Ver. 1.1, GAO-03-584G (2003).

⁹ See Timothy Sprehe and Charles McClure, "Study of Exemplary Practices in Electronic Records Management" (General Accounting Office, May 2003) and Karen Strong, "What's Your ECRM Number?" (Managing Electronic Records Conference, Chicago, May 21, 2007). Also see Christopher Becker, et al, "Assessing Digital Preservation Capabilities Using a Checklist Assessment Method," iPRES 2012, October 1 - 5, Toronto, Canada. Available at <https://ipres.ischool.utoronto.ca>

¹⁰ Tessella recently published a brief Digital Archiving Maturity Model White Paper. Available at <http://www.digital-preservation.com/wp-content/uploads/Maturity-Model-Web.pdf>.

¹¹ The genesis of this Digital Preservation Capability Maturity Model is rooted in a presentation given to the Arizona Electronic Records Management Task Force in 2002. The actual Digital Preservation Readiness Capability Maturity Model presented here was inspired in part by material developed by the International Records Management Trust to support assessment of the readiness of an organization to undertake an electronic records management program.

¹² Based on a 2010 review by the authors of the digital preservation practices of fifteen national, state and provincial-level programs.

¹³ The Council of State Archivists (www.statearchivists.com) is a national organization comprising the individuals who serve as directors of the principal archival agencies in each state and territorial government. Under regulations of the National Historical Publications and Records Commission, these individuals also serve as the State Historical Records Coordinators who chair their respective State Historical Records Advisory Boards (SHRABs).

¹⁴ CoSA SERI Phase 1, Mapping of Survey Results to the Digital Preservation Capability Maturity Model – Findings and Recommendations, Charles M. Dollar and Lori J. Ashley, September 2011, page 4.

¹⁵ SIO is a section of the International Council of Archives (www.ica.org).

¹⁶ The founding members of www.savingthedigitalworld are Lori Ashley, Charles Dollar, Michael Peterson, Bob Rogers, and the late Don Post.