Digital Preservation Capability Maturity Model© (DPCMM)

BACKGROUND AND PERFORMANCE METRICS

Version 2.6 Released May 16, 2014

This document provides an overview of the Digital Preservation Capability Maturity Model© (DPCMM) including its origins and foundations, performance metrics, and suggested use. The purpose of DPCMM is to provide practitioners with a process model and strategic planning tool to aid in benchmarking and improving digital preservation capabilities.

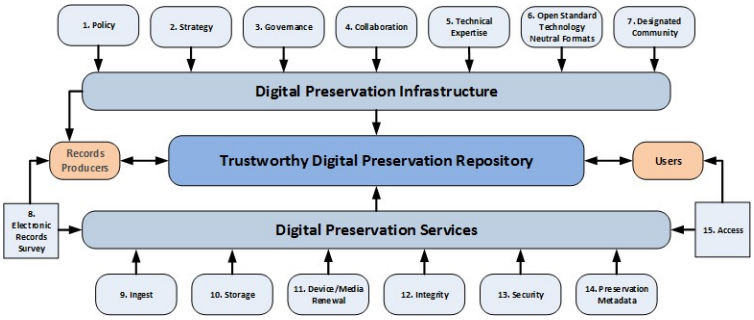


Table of Contents

Introduction

A common lament from the archives, library, records management, and business communities is that ensuring access to authentic, usable electronic records that have long-term[1] operational, regulatory, legal, or cultural memory value is so complex, perplexing, and costly it is difficult to know where and how to begin. This is a message that we understand and appreciate because we have heard it from our clients and from information, records and archives management practitioners in the public, private and not-for-profit sectors. Nonetheless, we believe that it is possible to deconstruct aspects of digital preservation to a level that is readily understood and enable dialogue and planning for implementation of a digital continuity[2] program within the available resources of most organizations.

We developed a Digital Preservation Capability Maturity Model© (DPCMM) that can be used to conduct a gap analysis of your organization's current capabilities and to delineate a multi-year roadmap of incremental improvements. The DPCMM, which is described in some detail in this document, draws upon functions and preservation services identified in ISO 14721, the Open archival information systems ("OAIS") Reference Model, as well as attributes specified in ISO 16363, Audit and Certification of Trustworthy Repositories. It is important to note that the DPCMM is not a "one size fits all" approach to ensuring long-term access to authentic electronic records. Rather, it is a flexible approach that can be adapted to each organization's specific requirements and resources, and takes into account a range of potential tools, systems, and implementation strategies.

The DPCMM is used to identify core digital continuity requirements which form the basis for debate and dialogue regarding the desired future state of each organization's digital preservation capabilities and the level of risk its leadership is willing to take on with regard to its electronic records. In many instances, this is likely to come down to the question of what constitutes digital preservation capability that is "good enough" to fulfill the organization's mission and meet the expectations of its stakeholders.

We hope that you find the model useful. We welcome your feedback. Please contact us via email or by visiting www.securelyrooted.com. Thank you.

Charles M. Dollar
thecdollar@att.net

Lori J. Ashley
loriashley@wi.rr.com

[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 greater (10+ years).

[2] Digital continuity refers to the ability of an organization to ensure digital information is accessible and usable by those who need it for as long as it is needed.

Strategic Planning for Long-Term Preservation of Valued Information Assets

The requirement and duty to preserve electronic records and other digital information assets should be driven by an organization's mission, vision, values and guiding principles. In our experience, every organization – regardless of size or sector – now has permanent and long-term records that exist in digital formats. Long-term digital continuity requires planning, resources, commitment and the ability to adapt to ever-changing operational, legal, regulatory, economic, social, and technology environments and requirements.

Strategic digital preservation planning should be based on rigorous external and internal environmental scans. External scans of the environment will identify laws and regulations, standards, best practices, and benchmarks that bear upon long-term access to and protection of electronic records. Requirements identified from external environmental scans of this domain have been collected and are available on various public sector and industry websites. A robust set of resources in the form of national and international standards, specifications, protocols, and tools have emerged and should be factored into digital continuity planning exercises and repository service level agreements.

Internal environmental scans can be more challenging because they require knowledge about current and planned information technology systems and platforms as well as the file formats used to create and store electronic records. This includes an understanding of the functionality of business applications and document and content management systems that may be required to transfer digital content to a trustworthy digital preservation repository.

Based on each organization's specific goals and objectives, a strategy to achieve the desired future state of digital preservation capabilities and lifecycle control of its information assets can be developed that takes into account both the external and internal requirements and operating environments. A clear delineation of the roles and responsibilities across the chain of electronic records management for record producers, owners and custodians is a critical component of supporting organization-wide information governance.

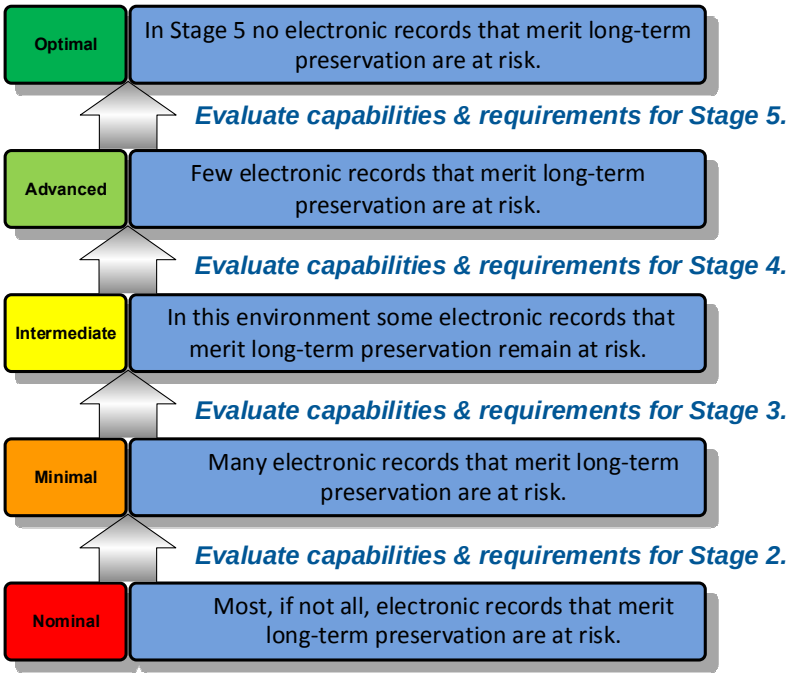
The planning framework references an incremental digital preservation capability improvement roadmap over a specified period of time. The level of improvement is likely to be shaped by an assessment of "at risk" digital information sets as well as available financial, technology and skilled human resources. The goal of strategic digital preservation planning is to establish and sustain one or more digital preservation repositories that conform to the specifications of ISO 14721, the de facto standard. In most instances, however, it is unlikely that 100 percent conformance with the specifications will ever be achieved. Prioritizing digital assets, repositories and transfer needs will help to rationalize investment decisions and engage business partners.

Strategic digital preservation planning is an on-going iterative process with no fixed end in sight, largely because changes in information technology and business requirements will introduce issues and concerns that require consideration of evolving tools, techniques, and approaches. Annual budgets and short-term tactical plans will have to be developed to support the incremental improvements envisioned in the plan. This iterative process should include periodic self-audits using standard auditing criteria (e.g., TRAC and ISO 16363) and checklists. Organizations should consider expanding their self-assessment process to include peer review and/or external third party auditing services as they become available[3].

Stages of the Digital Preservation Capability Maturity Model

A maturity model is a set of structured levels that describe how well the practices, processes and behavior of an organization can reliability and sustainably produce desired outcomes. The Digital Preservation Capability Maturity Model© (DPCMM)[4] is a five level (or stage) maturity continuum.

In Stage 1 a systematic electronic records management and/or digital preservation program has not yet been undertaken or a digital preservation program exists only on paper. The highest level (Stage 5) represents sustainable digital preservation capabilities systematically managed by process optimization and continuous process improvement. A high level description of key characteristics of each stage is provided on the next few pages.



[3] The ISO 16363 standard on auditing and certification criteria was released in 2012 and has been used for a series of internal and test audits. The framework for an international accreditation and certification process is under development.

[4] 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. Introduction to the potential use of an electronic records management capability maturity model by Timothy Sprehe and Charles McClure led to significant enhancements. The Digital Preservation Capability Maturity Model performance metrics were inspired in part by material developed by the International Records Management Trust to support an assessment of an organization's readiness to undertake an electronic records management program. The first use of DPCMM was in a 2007 project at the Delaware Public Archives.

The DPCMM is based on the functional specifications of ISO 14721, the auditing criteria of TRAC and ISO 16363, and accepted best practices in operational digital preservation repositories. It is a systems-based tool for charting an evolutionary path from disorganized and undisciplined management of electronic records, or the lack of a systematic digital continuity approach, into increasingly mature stages of digital preservation capability.

The goal of the Digital Preservation Capability Maturity Model (DPCMM) presented here is to identify at a high level where an electronic records management program is in relation to optimal digital preservation capabilities; report gaps, capability levels, and preservation performance metrics[5] to resource allocators and other stakeholders; and establish priorities for achieving enhanced capabilities to preserve and ensure access to long-term electronic records.

Stage 5: Optimal Digital Preservation Capability

Stage 5 is the highest level of digital preservation readiness capability that an organization can achieve. It includes a strategic focus on digital preservation outcomes by continuously improving the manner in which electronic records lifecycle management is executed. Stage 5 digital preservation capability also involves benchmarking infrastructure and services relative to other "best in class" digital preservation programs and conducting proactive monitoring for breakthrough technologies that can enable the program to improve its digital preservation performance. **In Stage 5 few if any electronic records that merit long-term preservation are at risk.**

Stage 4: Advanced Digital Preservation Capability

Stage 4 capability is characterized by an organization with a robust infrastructure and digital preservation services that are based on ISO 14721 specifications and TRAC, the Trustworthy Repository Audit and Certification: Criteria and Checklist and/or ISO 16363. At this stage the preservation of electronic records is framed entirely within a collaborative environment in which there are multiple participating stakeholders. Lessons learned from this collaborative framework serve as the basis for adapting and improving capabilities to identify and proactively bring long-term electronic records under lifecycle control and management. **Some electronic records that merit long-term preservation may still be at risk.**

[5] The performance metrics were applied while using DPCMM in consulting projects and underwent a significant revision in conjunction with a project sponsored by the Council of State Archivists (CoSA) to adapt the model to a digital preservation capability web survey for fifty-six state and territorial archives. Gary Miller (Wind Lake Solutions), Richard Pearce-Moses (Clayton State University), Milovan Misic (World Intellectual Property Organization) and Ton Bezemer (Anth.P.Bezemer LLM, The Netherlands) provided valuable commentary during development of the CoSA Digital Preservation Capability (DPC) Self-Assessment.

Stage 3: Intermediate Digital Preservation Capability

Stage 3 describes an environment that embraces the ISO 14721 specifications and other best practice standards and schemas and thereby establishes the foundation for sustaining enhanced digital preservation capabilities over time. This foundation includes successfully completing repeatable projects and outcomes that support enterprise digital preservation capabilities and fosters collaboration, including shared resources, between record producing units and entities responsible for managing and maintaining trusted digital repositories. In this environment **many electronic records that merit long-term preservation are likely to remain at risk.**

Stage 2: Minimal Digital Preservation Capability

Stage 2 describes an environment where an ISO 14721-based preservation repository is not yet in place. A surrogate preservation repository[6] for electronic records is available to some records producers that satisfies some but not all of the ISO 14721 specifications. There is some understanding of digital preservation issues and strategies but it is limited to a relatively few individuals. There may be virtually no relationship between the success or failure of one digital preservation initiative and the success or failure of another one. Success is largely the result of exceptional (perhaps even heroic) actions of an individual or a project team. Knowledge about such success is not widely shared or institutionalized. **Most electronic records that merit long-term preservation are at risk.**

Stage 1: Nominal Digital Preservation Capability

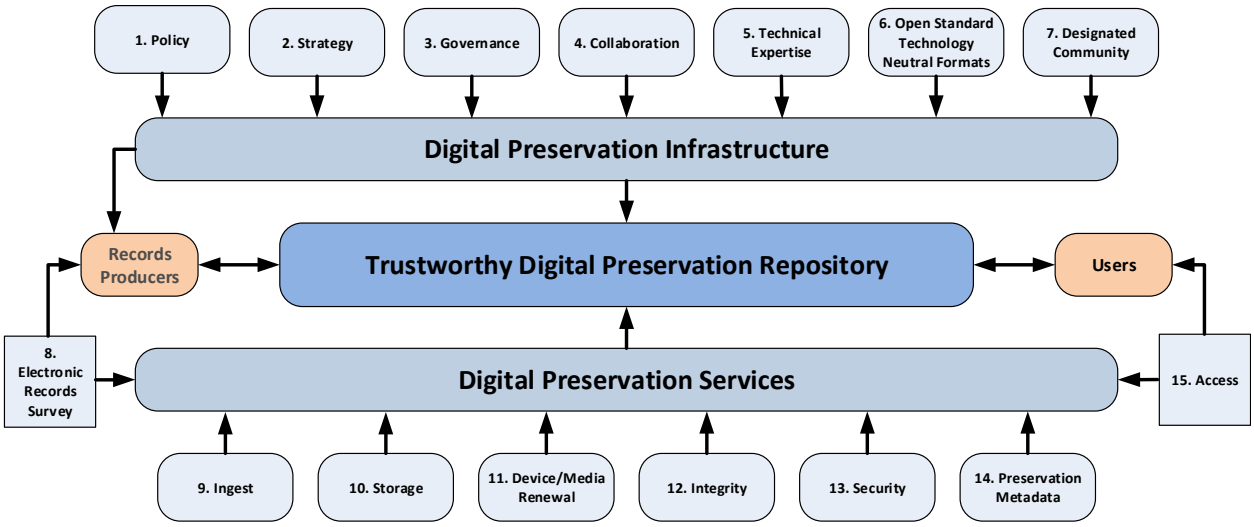
Stage 1 describes an environment in which the specifications of ISO 14721 and other standards may be known, accepted in principle, or under consideration but have not been formally adopted or implemented by the unit responsible for preservation (i.e., the Archives or Records Management function) or by Records Producers. Generally, there may be some understanding of digital preservation issues and concerns but this understanding is likely to consist of ad hoc electronic records management practices and digital continuity infrastructure and initiatives. Although there may be some isolated instances of individuals attempting to preserve electronic records on a network or removable storage media (e.g., DVD or hard drive), **practically all electronic records that merit long-term preservation are at risk.**

[6] A surrogate repository is a repository used for the preservation of long-term or permanent electronic records that does not fully conform to the specifications in the ISO 14721 reference model.

Scope of the Digital Preservation Capability Maturity Model

This capability maturity model consists of fifteen (15) components, or key process areas[7], that are necessary and required for the long-term continuity, access, and preservation of authentic, accessible and reliable electronic records. Each component is described and metrics for each of the five (5) levels of digital preservation capability are identified.

The objective of the model is to provide a process and performance framework (benchmark) against best practice standards and foundational principles of records management, information governance, and archival science. Note that the model has three separate but interrelated high level features: Digital Preservation Infrastructure, Trustworthy Preservation Repository, and Digital Preservation Services. Scope notes for the graphic elements in the diagram are found below.



Records Producers Records creators and owners who have an obligation to preserve records stored in digital format. These stakeholders either have the obligation or the option to transfer permanent and long-term (10+ year retention) electronic records to one or more specified preservation repositories for safekeeping and access.

Users Individuals or groups that have an interest in and/or right to access records held in the preservation repository. These stakeholders represent a variety of interests and access requirements that are likely to change over time.

[7] A Key Process Area (KPA) identifies a set of related activities that when performed together achieve a set of goals considered important.

Digital Preservation Infrastructure There are seven (7) infrastructure components that are essential to ensure a sustained commitment, including adequate resources, to the long-term preservation of electronic records:

1. Digital Preservation Policy
2. Digital Preservation Strategy
3. Governance
4. Collaboration
5. Technical Expertise
6. Open Standard Technology Neutral ("OS/TN") Formats
7. Designated Community

The seven digital preservation infrastructure components focus on what an organization as a distinct entity does to enable a preservation repository to execute the appropriate digital preservation services. Or to put it differently, a trusted preservation repository executes services within the constraints of the organization's digital preservation infrastructure.

Preservation Repository Ensuring the continuity of electronic records and enabling the design, operation, and management of preservation repositories requires the integration of people, processes, and technologies. The most complete preservation environment is based on models and performance criteria which include ISO 14721, ISO 16363, and generally accepted operational practices. The organization that owns the records may manage the repository or an external third party may provide this service.

A preservation repository may range from a simple system that involves a low-cost file server and software that provides non-integrated preservation services to complex systems comprised of data centers and server farms, computer hardware and software, and communication networks that are interoperable. It is likely that many organizations initially will rely on "surrogate" digital preservation capabilities and services that approximate but do not offer all of the transfer, data management and access functionality of an ISO 14721 conforming system.

Digital Preservation Services There are eight (8) key business process areas needed for continuous monitoring of external and internal environments in order to plan and take actions to sustain the integrity, security, usability and accessibility of electronic records stored in trustworthy preservation repositories:

8. Electronic Records Survey
9. Ingest
10. Archival Storage
11. Media/Device Renewal
12. Integrity
13. Security
14. Preservation Metadata
15. Access

The eight digital preservation services focus on a range of actions required to ingest and sustain long-term and permanent electronic records and continuously monitor the technical environment upon which they depend. The ability to plan actions to sustain the integrity, security, usability and accessibility of the records stored in the repository relies on the record producing organization systematically identifying and transferring electronic records of long-term value and providing sufficient strategic direction and resources.

Digital Preservation Capability Index Score

Each digital preservation capability metric has a value between 0 and 4. If a self-assessment or mapping exercise determines that a performance score of 3 for a specific component is appropriate, for Digital Preservation Policy, for example, it becomes the **index value** for the organization's Digital Preservation Policy capability. This procedure is repeated for the remaining fourteen elements of the DPCMM which results in an aggregated digital preservation capability score.

The range of composite index scores organized by each of the five levels is:

Capability Levels	Composite Index Score
Nominal Digital Preservation Capability	0
Minimal Digital Preservation Capability	1 - 15
Intermediate Digital Preservation Capability	16 - 30
Advanced Digital Preservation Capability	31 - 45
Optimum Digital Preservation Capability	46 – 60

Surrogate Infrastructure and Preservation Services Capabilities

The DPCMM is based on ISO 14721 and audit criteria (TRAC and ISO 16363) which set a high threshold for digital preservation capabilites. It is likely that many organizations with a mandate to preserve and provide access to long-term and permanent electronic records do not yet have the expertise and resources to implement a preservation repository that conforms to the ISO 14721 specifications and best practices. Most have not yet adapted their traditional records management and archival practices to fully address all of the demands of the digital information age and thus have significant gaps between their authority to preserve and disposition electronic records and their capabilities to fulfill these duties.

Some organizations have accessioned "born digital" or scanned records through manual or semi-automated workflows while others are addressing digital preservation requirements with a range of tools and services such as Contentdm®, ArchiveIt, BagIt, and LOCKSS, among others. Other institutions are participating in email management projects, leveraging technical capabilities of externally funded projects, or collaborating in funded projects. Despite not fully conforming with ISO 14721 specifications, these tools and projects are noteworthy and clearly are substantive and represent important emerging capabilities.

A number of institutions have built or are testing preservation repositories that are intended to conform with ISO 14721/ISO 16363 specifications. The DPCMM takes into account this spectrum of digital preservation capabilities by distinguishing between ISO 14721 conforming and partially conforming capabilities and services. We identify the latter as supporting "surrogate" digital preservation capabilities and services. Many of the projects, initiatives, tools and services listed in the table below are surrogates.

Category of Surrogate Capability	**Examples**[8]
Service Bureau	Web archiving service Archive-It
Curatorial practice-based and middleware assessment consortium	PeDALS (uses LOCKSS)
Best practice for a specific collection	GeoMAPP (for superseded geospatial data sets)
Centralized regional repository	Washington State MSPP
Software or packaging file format to facilitate the transfer of digital content	BagIt
Capture and registration functionality of a DoD 5015.2 certified document and records management application (RMA)	Open source options (Alfresco, Nuxeo) as well as a variety of proprietary solutions by major vendors including EMC, IBM, Hyland, OpenText, Oracle, Vignette and others are available
Digital asset management software and/or hosting solutions that facilitate transfers, indexing, storage, search, and web-based access	Contentdm® and OCLC Hosting Services

While some surrogate capabilities may be adequate for several years to meet requirements for long-term continuity, preservation and access to electronic records, other capabilities are short-term, project-based, and/or severely limited in scope. Successful adoption and integration of "lessons learned" after a project based on surrogate capabilities can be difficult and much less certain to deliver sustainable digital preservation capabilities and services.

[8] This table does not presume to provide an inclusive list of available digital preservation capabilities, services or solutions.

Thresholds for Digital Preservation Capabilities

The DPCMM was designed for organizations and repositories charged with the preservation of long-term and permanent electronic records to benchmark their capabilities against the specifications of ISO 14721, ISO 16363, and digital preservation community accepted practices based upon mapping these capabilities to DPCMM performance metrics. As noted earlier, DPCMM uses performance metrics to distinguish between capabilities that rise to the level of ISO 14721 conformance and digital preservation infrastructure and services that do not fully conform.

To enable those who use the DPCMM to differentiate between surrogate and ISO 14721 conforming capabilities, a series of surrogate performance threshold metrics is displayed below. In addition, Appendix B identifies the "essential properties" of conforming SIPs, AIPS, DIPs.

1. DIGITAL PRESERVATION POLICY

Threshold requirements for a digital preservation policy include:

- Specific statements about the long-term commitment of the organization to the preservation of digital information and records
- Documentation beyond a statutory mandate to preserve public records regardless of format
- Policy is related to other enterprise or organizational policies and practices

2. DIGITAL PRESERVATION STRATEGY

Threshold requirements for a digital preservation strategy include:

- Description of how the organization will acquire and preserve electronic records on behalf of its stakeholders
- Description of the organization's approach to file format obsolescence
- Description of the organization's approach to storage device and media obsolescence

3. GOVERNANCE

Threshold requirements for a digital preservation governance framework include:

- Assigned accountability and authority for operation and sustainability of preservation repositories for permanent or long-term records
- Identified roles and responsibilities for all digital preservation stakeholders

4. COLLABORATION

Threshold requirements for stakeholder-based collaborative engagement include:

- Stated commitment to promoting a collaborative framework to achieve strategic and operational goals for preservation repositories
- Participation in at least one internal or externally funded digital preservation project that involves a minimal level of peer collaboration

5. TECHNICAL EXPERTISE

Threshold requirements for digital preservation technical expertise include:

- Technical expertise in DoD 5015.2 conforming software within the organization's staff for internal use and/or to assist records producers
- Experience using surrogate digital preservation capabilities
- Experience transferring permanent electronic records to an external collaborative repository

6. OPEN STANDARD TECHNOLOGY NEUTRAL (OS/TN) FORMATS

Threshold requirements for capabilities related to Open Standard Technology Neutral Formats include:

- Adoption by the organization of at least one open standard technology neutral file format
- Active monitoring of the status of a minimal set of sustainable open standard technology neutral file formats through an external service like the US Library of Congress and leveraging the information for preservation planning purposes

7. DESIGNATED COMMUNITY

Threshold requirements for capabilities related to the Designated Community include:

- Agreements with selected records producers that enable the preservation repository to ingest surrogate SIPs
- Solicitation of input from selected users to identify their specific needs and requirements for access to electronic records in its custody
- On-going engagement with selected users that support dissemination of partially conforming DIPs by the preservation repository

8. ELECTRONIC RECORDS SURVEY

Threshold requirements for capabilities related to an electronic records survey of permanent records include:

- Analysis of existing retention schedules and collection of supplemental information about electronic records that will be transferred to the preservation repository
- Systematic analysis of records held by selected records producers to identify permanent electronic records at risk

9. INGEST

Threshold requirements for ingest capabilities of a preservation repository include:

- Accepts surrogate SIPs that contain descriptive content, context and structure information as well as access rights and restrictions
- Performs virus checks and format validations
- Produces surrogate SIPs for transfer to archival storage

10. ARCHIVAL STORAGE

Threshold requirements for archival storage capabilities of a preservation repository include:

- Validation of successful receipt of surrogate AIPs from the ingest process
- Validation of records deposited into archival storage through the use of checksums and error logs
- Creation and maintenance of information that establishes the provenance (i.e., electronic chain of custody) of surrogate Archival Information Package (AIPs) and describes the outcomes of preservation activities
- Two or more tier levels are used in archival storage

11. DEVICE/MEDIA RENEWAL

Threshold requirements for device/media renewal used in a presrevationl repository include:

- Existing protocol recommends device and storage renewal when they are on the verge of becoming obsolescent
- Existing protocol and procedures that require device/media renewal on a regularly scheduled basis

12. INTEGRITY

Threshold requirements for ensuring integrity of electronic records in a preservation repository include:

- Generation of MD5 hash digests for all internal data transmissions as well as before and after device and media renewal that are stored with partially conforming AIPs
- Generation of SHA-1 hash digests for all internal data transfers as well as before and after device and media renewal that are stored in partially conforming AIPs

13. SECURITY

Threshold requirements for ensuring security of electronic records in a preservation repository include:

- Disaster recovery procedures, one back-up copy of electronic records, and firewalls support the repository
- Role-based access rights limit access to authorized users

14. PRESERVATION METADATA

Threshold requirements for Preservation Metadata in a preservation repository include:

- A limited preservation metadata schema supports a minimal chain of custody
- A preservation metadata schema supports a robust chain of custody

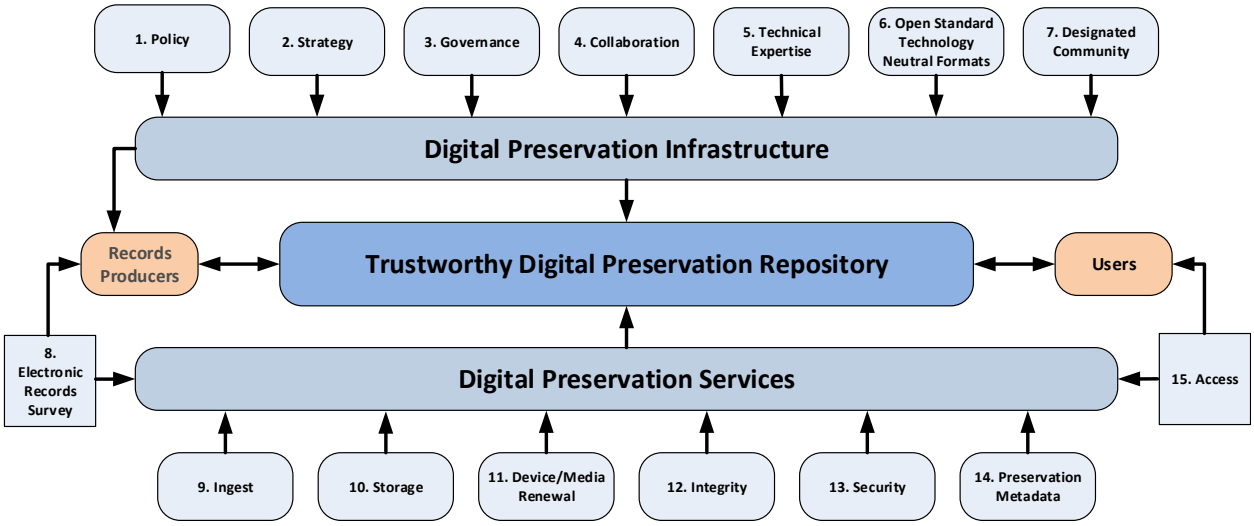
15. ACCESS

Threshold requirements for access to electronic records in a preservation repository include:

- Production of partially-conforming DIPs
- Dissemination of electronic records in only a single format (e.g., digital photographs)
- Dissemination of DIPs in two formats (e.g., PDF/A, JPEG, and TIFF

Digital Preservation Capability Metrics

In the next section the fifteen components of the DPCMM are described and capability metrics for each stage (or level) are provided.



NOTE: Throughout the metrics the term "organization" is used to denote the entity charged with the continuity and preservation of long-term and/or permanent records. The term "preservation repository" is used to denote the integrated people, processes and technologies charged with ingesting, storing, protecting, managing and providing access to the electronic records.

We recognize and acknowledge that different units and/or stakeholders may have these responsibilities. We encourage the use of alternative terms for these conventions.

Digital Preservation Infrastructure Components

1. Digital Preservation Policy

Preservation of accessible, authentic, and usable electronic records for as far into the future as necessary relies on digital information technologies but is equally dependent upon organizational commitment and practices. The organization charged with ensuring preservation and access to long-term and permanent legal, fiscal, and/or historical records should state its policy in writing, communicate the policy to all stakeholders, and periodically audit the policy for compliance.

A written digital preservation policy includes the purpose, scope, accountability, and approach to the transfer of records and the operational management and sustainability of trustworthy preservation repositories.

Level	Digital Preservation Policy Capability Metrics
0	The organization does not have a written digital preservation policy.
1	The organization has a digital preservation policy in development but it has not yet been approved or issued.
2	The organization has issued a digital preservation policy and it is widely disseminated to stakeholders.
	ISO 14721 Conformance
3	The organization annually conducts a self-assessment and reports adherence to the digital preservation policy to its governing body.
4	The organization arranges for a periodic peer review or external audit of the digital preservation policy and revises the policy as appropriate.

2. Digital Preservation Strategy

The organization charged with the preservation of long-term and permanent electronic records must proactively address risks associated with technology obsolescence. While no single strategy is appropriate for all organizations, data types and resources, there must be plans to periodically upgrade storage devices, storage media, and file formats.

Left unchecked the obsolescence of storage devices and media eventually will render the bit streams of electronic records unreadable. The inevitable obsolescence of file formats, especially native, proprietary ones, means that over time software applications will not be able to render bit streams into understandable and usable electronic records.

The generally accepted strategy is to mitigate the obsolescence of storage devices/media through planned, periodic renewal, which over time ensures that "bit streams" can be read by current technologies (*see Component 11*). The generally accepted strategy for mitigation of file format obsolescence is reliance on interoperable, open standard technology neutral formats, which are otherwise considered "preferred preservation formats" (*see Component 5*).

Level	Digital Preservation Strategy Capability Metrics
0	The organization does not have a formal strategy to address technology obsolescence.
1	The strategy calls for accepting electronic records in native formats on an ad hoc basis and keeping the bit streams alive until software and other resources are available to transform the records into open standard technology neutral file formats.
2	The strategy calls for encouraging Records Producers to convert electronic records of long-term and permanent value in their custody to "preservation ready" formats at or near the time of receipt and creation. The strategy includes ad hoc monitoring of changes in technologies that may impact digital records collections in the custody of Records Producers and preservation repositories.
	ISO 14721 Conformance
3	The strategy calls for transformation of electronic records in selected native file formats to preferred preservation formats at ingest and proactive monitoring of changes in technologies that affect the preservation of electronic records.
4	The strategy calls for the transformation of all electronic records in native file formats to preferred preservation formats at ingest. Electronic records in archival storage are automatically transformed to newer interoperable forms as they displace current ones.

3. Governance

An organization with a digital preservation mandate should have a formal decision-making process aligned to its enterprise information governance[9] framework that assigns accountability and authority for the preservation of electronic records with permanent value, and articulates approaches and practices for preservation repositories sufficient to meet stakeholder needs. This capability ideally leverages existing organizational rules, practices and protocols as well as engages cross-functional stakeholders.

Long-term preservation, however, may require the creation of new authorities to address the threats of technology obsolescence. A preservation repository may be run by a business or technology unit, operated as one or more standalone repositories under the control of a Records Management unit or Archives, include participation in a federated or regional repository system, and/or use digital preservation services provided by one or more third parties.

The organization exercises digital preservation governance in conjunction with archives, information management/technology functions, and with other custodians and digital preservation stakeholders such as Records Producers and Users. The governance framework enables compliance of the preservation repository with applicable laws, regulations, record retention schedules, disposition authorities, and standards. Plans and decisions resulting from governance activities, including repository operational statistics, are shared with internal stakeholders and third party operators.

Level	Governance Capability Metrics
0	The organization's current information governance activities do not specifically address digital preservation requirements.
1	The organization has a limited, project-based digital preservation governance framework that is operational or has been successfully completed.
2	The organization is developing an enterprise governance framework that identifies roles and responsibilities for electronic records life cycle management and digital preservation.
	ISO 14721 Conformance
3	The organization has adopted an enterprise digital preservation governance framework that includes comprehensive policies and procedures and specifies an on-going commitment to the sustainability of one or more preservation repositories.
4	The enterprise digital preservation governance framework supports one or more preservation repositories and is reviewed and updated at least every two years to take into account changing technologies and organizational requirements.

[9]Gartner's definition: ***Information governance*** is the specification of decision rights and an accountability framework to encourage desirable behavior in the valuation, creation, storage, use, archival and deletion of information. It includes the processes, roles, standards and metrics that ensure the effective and efficient use of information in enabling an organization to achieve its goals.

4. Collaboration

Digital preservation is a multi-faceted discipline that takes into account the organization's information architecture and technology environment as well as accepted standards and best practices. An organization with a mandate to preserve electronic records is well served by maintaining and promoting collaboration among its many stakeholders.

Plans for different types of records, models for preservation approaches and criteria, and a framework of repository components and services require tighter cooperation and engagement between long-standing partners such as IT, peer organizations, software and service providers, and other support functions. Collaboration should acknowledge the interdependencies between and among the operations of Records Producers, legal and statutory requirements, information technology policies and governance, and historical accountability.

Active engagement in addressing the challenges of long-term digital preservation makes the best use of resources and lessons learned. The collaborative framework evolves in response to changes in information technologies and the business operations of Record Producers. This collaborative framework seeks to leverage financial, human, and technical resources, promote stewardship, and exchange knowledge about the current and future state of digital initiatives. This collaborative framework may extend beyond the organization to include other repositories, federal or other public sector agencies, as well as consortia of other organizations with a similar or shared mission.

Level	Collaboration Capability Metrics
0	No collaborative digital preservation environment exists within or across the organization.
1	The organization is currently working to establish a framework for collaborative engagement on electronic records management and digital preservation issues.
2	Under its collaborative digital preservation framework the organization has successfully engaged or is currently engaged with selected stakeholder entities to proactively address digital preservation requirements. These engagements may include externally funded collaborative digital preservation initiatives.
	ISO 14721 Conformance
3	Under its collaborative digital preservation framework the organization has successfully engaged or is currently engaged with most stakeholders to proactively identify and meet their digital preservation requirements.
4	The organization continuously monitors and updates its digital preservation collaboration framework to support proactive outreach to all stakeholders to identify and meet their digital preservation requirements.

5. Technical Expertise

A viable digital preservation capability requires organizations to have sufficient expertise in electronic records management and digital preservation to support all of the infrastructure and requisite key processes, including on-going professional development for personnel and certification of the repository. Technical expertise may exist within internal or contracted staff, may be provided by a centralized service bureau, or by external service providers.

NOTE: It is likely that many organizations will initiate a long-term digital preservation program with one or more electronic records management applications (RMA) that conform with country or regional-level standards, such as Department of Defense (DoD) Directive 5015.2-STD, Model Requirements for the Management of Electronic Records (MoReq2010), and Victorian Electronic Records Strategy version 2 (VERS2). These systems support some features of ISO 14721.

Level	Technical Expertise Capability Metrics
0	The organization has little or no operational access to specialized professional technical expertise in digital preservation or electronic records management.
1	The organization has access to internal or external professional technical expertise that supports only narrowly defined project-based digital preservation initiatives. This may also include technical expertise in deploying electronic records management applications (RMA) certified to one or more standards.
2	The organization has access to internal or external professional technical expertise who assist Records Producers in the creation of preservation-ready records and/or support surrogate ingest and archival storage services.
	ISO 14721 Conformance
3	The organization has access to internal or external professional technical expertise that supports all functions of an ISO 14721 preservation repository.
4	The organization has access to internal or external professional technical expertise that supports all functions of an ISO 14721 preservation repository, along with the capability to assess the impact of emerging technologies that should be taken into account in long-term digital preservation planning activities.

6. Open Standard Technology Neutral Formats

A requisite for a sustainable digital preservation program that ensures long-term access to usable and understandable electronic records is mitigation of file format obsolescence. Current best practice for mitigation of file format obsolescence involves three separate but related actions.

> The first action is to support a Technology Watch Program on the sustainability of file formats. This can be achieved through an external service like the U.S. Library of Congress[10] or PRONOM[11], the technical registry of the National Archives of the United Kingdom.
>
> The second action involves the commitment of the preservation repository to adopt open standard technology neutral ("OS/TN") file formats to use as preservation formats.
>
> The third action pertains to proactive engagement and collaborative working relationships with Records Producers to advise them on the use of preservation-ready file formats when they create and maintain electronic records of long-term and permanent historical, legal, or financial value that will be transferred to the custody of a preservation repository.

Open standard platform-neutral file are developed in an open, public setting, are issued by a certified standards organization, and have few or no technology dependencies. Current preferred OS/TN file formats include:

> CSV for spreadsheets
> HTML, Plain Text, XML, ODF, and PDF/A for text
> JPGE 2000 for photographs
> PDF/A, PNG, and TIFF for scanned images
> SVG for graphics
> MPEG-4 and Motion JPEG2000 for video
> WAVE_BWF LPCM for audio
> WARC for web pages

Over time digital preservation tools and solutions will emerge that will require new open standard technology neutral standard file formats. Open standard technology neutral formats are backwardly compatible so they can support interoperability across technology platforms over an extended period of time and space.

Capability metrics for Open Standard Technology Neutral Formats are provided on the next page.

[10] Visit the Library of Congress Digital Formats Web Site at www.digital preservation.gov/formats/index.shtml
[11] Visit http://www.nationalarchives.gov.uk/PRONOM/Default.aspx

Level	Open Standard Technology Neutral Formats Capability Metrics
0	The organization has not yet adopted any open standard technology (OS/TN) file format as a preferred preservation format.
1	The organization has adopted at least one OS/TN file format as a preferred preservation format.
2	The organization has adopted at least three OS/TN formats as preferred preservation formats.
	ISO 14721 Conformance
3	The organization has adopted at least five open standard technology neutral formats as preferred digital preservation formats (text, spreadsheets, scanned images, vector graphics, digital photos, audio, video, and web pages). A Technology Watch Program is used to monitor the sustainability of these OS/TN file formats.
4	The organization has adopted at least ten OS/TN neutral formats as preferred digital preservation formats and continuously monitors the emergence of new OS/TN file formats and adopts them as appropriate for use as preferred digital preservation formats.

7. Designated Community

The organization that has responsibility for preservation and access to permanent electronic records is well served through proactive outreach and engagement with its Designated Community of Records Producers and Users. While this activity has traditionally taken place with representatives of the Records Producers in the form of records appraisal and retention schedule review and disposition authorization, the challenges of digital preservation demand that records management practitioners engage in additional "upstream" actions in the lifecycle management of long-term and permanent electronic records. Submission agreements[12] and transfer protocols should be standardized and service level agreements defined for repository operations. Formal agreements and procedures with Records Producers document the content, rights, and conditions under which the preservation repository will ingest, preserve, and provide access to electronic records. Specific assurances are given to ensure privacy and protection of intellectual property as appropriate.

The organization maintains written procedures regarding access to its electronic collections. Dissemination Information Packages (DIPs) are developed and updated in conjunction with its User communities (e.g., scholars, genealogists, the public, etc.). Procedures are regularly reviewed and updated to take into account changing business practices of Records Producers as well as the research interests and access capabilities of Users.

Level	Designated Community Capability Metrics
0	The organization has no formal documentation that defines the rights, obligations, and responsibilities of the Designated Community for electronic records to be transferred to or held by a preservation repository.
1	The organization has ad hoc agreements with selected Records Producers that support the transfer of electronic records to a preservation repository.
2	The organization has formal, written agreements with a few Records Producers that support the transfer of surrogate SIPs and proactively reaches out to select Users to identify their specific needs and requirements for access to electronic records in its custody.
	ISO 14721 Conformance
3	The organization works with most Records Producers in its mandated domain to establish formal agreements about their rights, obligations, and responsibilities for transferring Submission Information Packages (SIPs) to the preservation repository. The organization works closely with most Users to establish DIP profiles that meet their needs and requirements.
4	The organization actively engages all Records Producers in its mandated domain to establish written agreements about their rights, obligations and responsibilities for transferring SIPs. Profiles of conforming SIPS are regularly reviewed and updated to take into account changing business practices of Records Producers. The organization works closely with all Users to establish DIP profiles that meet their evolving needs and requirements.

[12] Submission agreements specify the data model and the logical constructs used by the Records Producer and how they are represented on each media delivery to the repository.

8. Electronic Records Survey

All public and private organizations are responsible for records created, received or acquired that are evidence of its business activities, regardless of the format or media used. They have an obligation to ensure the authenticity, integrity, usability and reliability of the records for as long as they are required.

Records with long-term retention requirements or permanent value are often transferred from decentralized operations 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. One effective way to accomplish this is for the organization to maintain a comprehensive inventory of electronic records and systems as well as collaborative working relationships between stakeholders that include Records Producers, Legal/Compliance, Archives, Records Management, Information Services/Technology and third party application, solution and service providers.

A key feature of conforming ISO 14721 digital repositories is reliance on open standard technology neutral formats. During the ingest process electronic records in proprietary formats are transformed into preferred preservation formats that the organization and/or repository has 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 electronic records with retention requirements of ten (10) years or more in order to support planning and preservation activities:

- "Preservation Ready" electronic records.
- "Near-Preservation Ready" records, that is electronic records in formats for which tools are available that can export native format documents to open standard interoperable technology neutral formats. An example is Microsoft Word 2007 that contains a tool to transform Word documents into PDF/A format.
- “Legacy” records, that is, electronic records in a proprietary native format for which no export 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 Records Producers, interviews with selected business units or third parties that routinely create, receive or acquire electronic records, review of records retention and disposition schedules, analysis of the organization’s information technology portfolio, as well as the use of search engines and algorithms to identify specific file formats currently used in the capture and storage of electronic records on network drives.

Capability descriptions for Electronic Records Survey are provided on the next page.

Level	Electronic Records Survey Capability Metrics
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 uses existing retention schedules to identify electronic records of permanent historical, fiscal, and legal value in the custody of Records Producers. 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 select Records Producers. This effort may be enhanced by focusing on identified "at risk" electronic records.
	ISO 14721 Conformance
3	The organization supplements analysis of "at risk" electronic records through collection of information about the volume and location, media and format types (preservation ready and near-preservation ready) of permanent electronic records in the custody of Records Producers.
4	The organization has identified and categorized all preservation ready, near-preservation ready, and legacy permanent electronic records in the custody of all Records Producers.

Digital Preservation Services Components

9. Ingest

A preservation repository that conforms to ISO 14721 functional specifications and associated best practices has the capability to systematically ingest (receive and accept) electronic records from Records Producers in the form of Submission Information Packages (SIPs).

The preservation repository accepts SIPs from Records Producers, validates the agreements and integrity of the digital content, moves the SIPs to a staging area where virus checks and content and format validations are performed, transforms electronic records into designated preservation formats as appropriate, extracts metadata from SIPs and writes it to Preservation Description Information (PDI), creates Archival Information Packages (AIPs), and transfers the AIPs to the repository's storage function.

Level	Ingest Capability Metrics
0	The organization does not have a digital preservation repository capable of receiving or ingesting long-term and permanent electronic records.
1	The preservation repository receives electronic records from Records Producers based on ad hoc agreements without regard to format, integrity, virus checks, and metadata quality. None of this rises to the level of an ISO 14721 conforming SIP.
2	The repository receives surrogate SIPs that are held in a staging area while virus checks and format validations are manually executed. Surrogate AIPs are manually created and transferred to archival storage.
	ISO 14721 Conformance
3	The preservation repository ingests SIPs through semi-automated means that validate Record Producer agreements and integrity of the digital content. SIPs are moved to a staging area where virus checks and content and format validations are performed and electronic records transformed into designated preservation formats as appropriate. Metadata is extracted from SIPs and written to Preservation Description Information (PDI). Archival Information Packages (AIPs) are created and transferred to the repository's storage function.
4	The preservation repository ingests SIPs through automated means that validate Record Producer agreements and integrity of the digital content. SIPs are moved to a staging area where virus checks and content and format validations are performed and electronic records transformed into designated preservation formats as appropriate. Metadata is extracted from SIPs and written to Preservation Description Information (PDI). Archival Information Packages (AIPs) are created and transferred to the repository's storage function.

10. Archival Storage

The ISO 14721 open archival information system reference model delineates a number of systematic automated storage services that support receipt and validation of successful transfer of AIPs from ingest, creation of Preservation Description Information (PDI) for each AIP that confirms its fixity (i.e., no corruption has occurred) during any preservation actions through the capture and maintenance of error logs, updates to PDI, including transformation (i.e., migration) of electronic records to new formats, three storage tiers (on-line, near-line, and off-line), production of Dissemination Information Packages (DIPs) for access, and collection of operational statistics.

Archival storage is dependent on other preservation services depicted in the capability maturity model including Device/Media Renewal, Integrity and Security protections, and on the availability and enforcement of Preservation Metadata standards.

Level	Archival Storage Capability Metrics
0	The preservation repository either does not accession electronic records or its holdings consist of primitive archival storage (e.g., a shared drive or CDs/DVDs) where it is available.
1	A single instance of a preservation repository supports the storage of surrogate AIPs with limited metadata that can be mapped to Preservation Description Information (PDI).
2	A single instance of a surrogate preservation repository supports the storage of surrogate AIPs. Manual capture of partial Preservation Description Information (PDI) establishes the provenance of the AIPs.
	ISO 14721 Conformance
3	A single instance of a preservation repository supports the storage of AIPs. Semi-automated tools capture Preservation Description Information (PDI) that establishes the provenance of AIPs (e.g., format transformations, device/media renewal, integrity checks, and access/rights protection).
4	Two or more geographically separated instances of a preservation repository support the storage of AIPs. Automated capture of Preservation Description Information (PDI) establishes the provenance of AIPs. Automated capture of preservation repository storage and operational statistics supports on-going comprehensive digital preservation planning.

11. Media/Device Renewal

There is no known digital device or storage medium that is invulnerable to decay and obsolescence. A foundational digital preservation capability for an organization that has the responsibility to preserve electronic records of long-term and permanent value is ensuring the readability of the bit streams underlying the electronic records. ISO 14721 specifies that a trustworthy digital repository's storage devices and storage media should be monitored and renewed ("replicate"/"repackage") periodically to ensure that the bit streams remain readable over time.

Decay of magnetic and optical storage media is inevitable. Accelerated aging tests predict that most magnetic and optical storage media have a life expectancy of 100 years or more if stored in a controlled environment. However, predicted media life expectancies of hundreds or even thousands of years is of little practical benefit because the fundament issue in device and storage media for a preservation repository is technology obsolescence. This is likely to occur when:

- There is a change in the physical form factor (e.g., from 10.5 inch reels of magnetic tape to tape cartridges)
- There is a change in the method of physically encoding information on the recording surface that makes it impossible to transfer electronic content from an obsolescent tape or disk drive to a contemporary one.
- A vendor decides to discontinue a product
- A legacy system or application is decommissioned without exporting electronic records to the new computing environment.

A preservation repository should support a robust device/storage media renewal program. Depending upon available resources, this renewal program can range from non-network storage devices/storage media, to local network based storage devices/storage media, to external third parties that provide storage services that include device/storage media renewal. Regardless of how and when device/storage media renewal occurs, a critical requirement is that a protocol is in place that mandates the capture and preservation of the results of periodic validation of the integrity of electronic records before and after completion of device/digital storage renewal.

Capability descriptions for Media/Device Renewal are provided on the next page.

Level	Media/Device Renewal Capability Metrics
0	The preservation repository has no formal device and media renewal protocol in force.
1	The preservation repository mandates device/media renewal when they are on the verge of becoming obsolescent.
2	The preservation repository mandates device/media renewal on a regularly scheduled basis (e.g., every ten years).
	ISO 14721 Conformance
3	The current device and media renewal program supports an annual media inspection program that identifies preservation repository storage media facing imminent catastrophic data loss and executes device/media renewal as appropriate.
4	The current device and media renewal program continuously monitors the potential loss of the readability of electronic records and automatically replaces devices/storage media and writes the records to new storage media as appropriate.

12. Integrity

A key capability in conforming ISO 14721 preservation repositories is ensuring the integrity ("fixity") of records in its custody. Accidental or intentional alterations can occur during device/media renewal, internal data transfers, and other preservation actions. One way to establish integrity is through the use of cryptographic hash digests that are digital fingerprints of electronic records in a SIP, an AIP or some aggregation of them.

A cryptographic hash digest computed before a digital preservation operation and after its completion will detect any changes, even down to a single bit. Hash digests are stored in Preservation Description Information (PDI) in AIPs where they can be reviewed to confirm that no changes occurred during device/medial renewal, internal data transfers, and other preservation actions, thereby supporting an unbroken chain of electronic custody. The strength of hash digests varies, the lowest being MD5 and the highest is SHA-2.

NOTE: In October 2012 the National Institute of Standards and Technology (NIST) selected the algorithm to be used in SHA-3. NIST released the draft specification in April 2014.[13]

Hash digests do not support the chain of electronic custody when the preservation action involves format transformation because the underlying bit streams of transformed digital records will not match the bit streams before they were transformed. However, this can be compensated for with the collection of information about all of the preservation actions undertaken with regard to AIPs and storing this information in AIP Preservation Description Information. Affixing a digital signature to AIPs encapsulated in XML after each preservation action also provides a strong electronic chain of custody.

Level	Integrity Capability Metrics
0	The preservation repository has no documented procedure for integrity protection of electronic records in its custody.
1	The preservation repository generates and preserves MD-5 hash digests of electronic records before and after device/media renewal and other archival storage preservation actions.
2	The preservation repository generates and preserves SHA-1 hash digests before and after device/media renewal and other internal preservation actions.
	ISO 14721 Conformance
3	The preservation repository generates and preserves SHA-2 hash digests before and after device/media renewal and other internal preservation actions for all AIPs and stores them in Preservation Description Information (PDI).
4	The preservation repository generates SHA-2 hash digests, encapsulates them in AIPs in XML and signs them with a digital signature. Integrity protection procedures are continuously evaluated and updated as new tools and approaches become available.

[13] http://csrc.nist.gov/publications/drafts/fips-202/fips_202_draft.pdf

13. Security

Digital preservation requires processes that restrict access to the physical repository where digital content is stored, ensure the security of electronic records through techniques that block unauthorized access, protect the confidentiality and privacy of records and intellectual property rights, support periodic backup of electronic records that are stored at offsite storage repositories, and support disaster recovery and business continuity.

Level	Security Capability Metrics
0	The preservation repository does not have formal disaster recovery, backups, or firewall procedures in place to protect the security of electronic records.
1	The preservation repository supports the security of electronic records in its custody through disaster recovery procedures.
2	The preservation repository supports the security of electronic records in its custody through a comprehensive firewall protection.
	ISO 14721 Conformance
3	The preservation repository supports the security of electronic records in its custody through comprehensive role based access rights management.
4	The preservation repository support the security of electronic records in its custody by continuously monitoring security protection processes and revising them in response to evolving technology capabilities and changing business requirements.

14. Preservation Metadata

A preservation repository collects and maintains metadata that describes preservation actions associated with custody of permanent electronic records. Preservation metadata includes an audit trail that documents preservation actions carried out, why and when they were performed, how they were carried out and with what results.

A current best practice is the use of a PREMIS-based preservation metadata schema for all permanent electronic records to support an electronic chain of custody that documents authenticity over time as preservation actions are executed. Capture of all related metadata, transfer of the metadata to any new formats/systems, and secure storage of metadata is critical. All of this associated metadata is stored in the Preservation Description Information (PDI) component of ISO 14721 AIPs.

Level	Preservation Metadata Capability Metrics
0	A primitive preservation repository has little or no preservation metadata for electronic records in its custody.
1	The preservation repository supports an ad hoc preservation metadata schema and establishes a minimal chain of custody for electronic records in its custody.
2	The preservation repository supports a surrogate PREMIS metadata schema for electronic records in its custody.
	ISO 14721 Conformance
3	The preservation repository supports a manual based PREMIS preservation metadata schema for most of the electronic records in its custody.
4	The preservation repository supports an automated PREMIS preservation metadata schema that extracts preservation metadata from SIPs for all of the electronic records in its custody and writes it to AIPs.

15. Access

The purpose of digital continuity is to ensure that usable, understandable, and trustworthy electronic records are accessible as far into the future as may be necessary, subject to any restrictions imposed by the Records Producers. Consequently, communities of Users should have access to Dissemination Information Packages (DIPs) derived from Archival Information Packages (AIPs) that a trustworthy digital repository properly preserves.

This access capability may include the creation and maintenance of User searchable retrieval metadata that can be queried to identify information of interest and disclosure free (redacted to protect privacy, confidentiality, and other rights where appropriate) Dissemination Information Packages (DIPs). In no instance will Users have direct access to Archival Information Packages (AIPs).

Level	Access Capability Metrics
0	The preservation repository either has no electronic records in its custody or has no capability to support access to electronic records in its custody.
1	The preservation repository supports access to electronic records in a single format (e.g., JPEG or PDF) while enforcing all access restrictions.
2	The preservation repository supports access to electronic records in at least three open standard technology neutral formats (e.g., PDF/A, JPEG, and TIFF formats) while enforcing all access restrictions.
	ISO 14721 Conformance
3	The preservation repository produces DIPs containing at least six (6) open standard technology neutral formats, tracks User query trends that are used to identify the need for updated accessibility tools, and captures auditable documentation of the production of DIPs.
4	The preservation repository disseminates DIPs containing records in any format that Users request, enables redaction of electronic records in its custody with access restrictions where appropriate, tracks user query trends, and captures auditable documentation of the production of DIPs.

APPENDIX A: Glossary of Terms

Access. The OAIS entity that contains the services and functions which make the archival information holdings and related services visible to Consumers.

Access Rights Information: The information that identifies the access restrictions pertaining to the Content Information, including the legal framework, licensing terms, and access control. It contains the access and distribution conditions stated within the Submission Agreement, related to both preservation (by the OAIS) and final usage (by the Consumer). It also includes the specifications for the application of rights enforcement measures.

Archival Information Package (AIP). An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an ISO 14721 (OAIS) based digital repository.

Authenticity. The degree to which a person, object, activity, or event is what it purports to be. An authentic record is one that can be demonstrated by evidence to be what it purports to be.

Born Digital. Refers to materials that originate in digital form.

Chain of Custody. A formal procedure that documents an information object (e.g., a record) as always being in the custody of an entity (person, system, organization and the like) legally responsible for maintaining the integrity of the object.

Comforming AIPs. *See Essential Properties*.
Comforming DIPs. *See Essential Properties.*
Comforming SIPs. *See Essential Properties.*

Consumer. The role played by those persons, or client systems, who interact with OAIS services to find preserved information of interest and to access that information in detail. This can include other OAISs, as well as internal OAIS persons or systems.

Content Information. The set of information that is the original target of preservation. It is an Information Object comprised of its Content Data Object and its Representation Information. An example of Content Information could be a single table of numbers representing, and understandable as, temperatures, but excluding the documentation that would explain its history and origin, how it relates to other observations, etc.

Cryptograph Hash Algorithm. A mathematical transformation of digital content without regard to its size that reduces it to a fix-length string (e.g., 160 bits), which is called a hash value (sometimes called a message digest, a digital fingerprint, a digest or a checksum). A cryptographic hash algorithm is relatively easy to reproduce from the original data but it is computationally infeasible to reproduce the original string of data from a hash digest. It is also computationally infeasible that two slightly different strings of data will have the same hash digest. In digital preservation cryptographic hash algorithms play an important role in validating the integrity of digital records by demonstrating that no changes have occurred over time.

Comma-Separate Value (CSV). A de facto standard for importing and exporting tabular data from spreadsheets and databases. The tabular data consists of rows of plain text (e.g., ASCII) in organized fields (columns) that are delimited by separate by comas, semicolons, or spaces. Rows are considered as data records, each of which has the same sequence of fields.

Compression. A technique to reduce the volume of bits of digital objects being transferred or stored that can be reconstructed at the time of rendering. Typically, compression is associated with digital images and audio and video digital content. There are two forms of compression, lossy and lossless. Lossy references a compression technique that permanently removes some bits that cannot be restored during decompression. Lossless denotes a compression technique that enables restoration of all of the bits during decompression.

Designated Community. An identified group of potential consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities.

Digital Signature. A cryptographic technique for creating a bit stream that can be affixed to a document (or any other digital object) and thereby attest to its authenticity. A digital signature includes a private key that is known only to its owner and a reciprocal public key that can be made available to anyone. A digital object signed with a private key can only be validated by its reciprocal public key. It is computationally infeasible for anyone to generate a valid digital signature that does not possess the private key. It is computationally infeasible to create a private key from a public key.

Disclosure-free. Associated with a copy of a record that contains no personally identifiable information (e.g., a Social Security Number) or otherwise restricted access information. *See Redacted.*

Dissemination Information Package (DIP). The Information Package, derived from one or more AIPs, received by the Consumer in response to a request to the ISO 14721 (OAIS) based digital repository.

DoD 5015.2 Criteria for Electronic Records Management Software Applications. A standard that specifies mandatory and optional baseline functional requirements for records management application software employed in the Department of Defense. DoD 5015.2 certification means that a records management software application has passed received formal certification that it conforms to these specifications. Since its introduction, it has become a de facto standard for electronic records management applications.

Dublin Core. An international standard (ISO 15836), Dublin Core defines metadata elements that describe and support on-line access to material. It consists of 15 elements: title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage, and rights.

Electronic Record. Any combination of text, graphics, data, audio, pictorial or other information representation in digital form set aside for future reference that is created, used, modified, stored, and retrieved by a computer application/system.

Encapsulation. A technique for placing digital records and associated metadata in a container or wrapper that can be manipulated or transmitted without regard what the wrapper contains. XML supports the use of Document Type Definitions in wrappers that separate logical structures (i.e. content structure) from their rendered physical representations.

Essential Properties (SIP, AIP, and DIP). Attributes of conforming ISO 14721 Information Packages and digital preservation community best practices. *For more information see Appendix B.*

Extensible Markup Language (XML). XML is a World Wide Web Consortium (W3C) standard for marking up text based documents that are interoperable. Interoperability is achieved through the assignment of tags (Document Type Definition) to the logical structure of text based digital content and the use of a Style Sheet for rendering the content into human readable form. A Document Type Definition assigns tags that define the logical and semantic structure of text based documents. In the context of DPCMM XML can be considered a "preferred preservation format."

Fixity of Information. The information which documents the authentication mechanisms and provides authentication keys to ensure that the Content Information object has not been altered in an undocumented manner.

Format. A wrapper for the 1s and 0s of bit streams that underlie electronic records. It specifies how the 1s and 0s are encoded and how they are to be interpreted. Typically, the extension to electronic content denotes the format used (e.g., TXT for ASCII Text, PDF for Portable File Format).

Format Validation. The process that identifies the format of electronic records and confirms that the format used conforms to its formal published specifications.

Hyper Text Markup Language (HTML). HTML is a mark-up (i.e., tags) language initially designed (1990) for creating interoperable text, image, and audio digital context for web browsers. In 2000 it became an international standard: ISO 15445:2000.

Information Package. The Content Information and associated Preservation Description Information which is needed to aid in the preservation of the Content Information. The Information Package has associated Packaging Information used to delimit and identify the Content Information and Preservation Description Information.

Information Governance. Decision rights and an accountability framework to encourage desirable behavior in the valuation, creation, storage, use, archival and deletion of information. Information governance includes processes, roles, standards and metrics that ensure effective and efficient use of information in enabling an organization to achieve its goals.

Ingest. The OAIS entity that contains the services and functions that accept Submission Information Packages from Producers, prepares Archival Information Packages for storage, and ensures that Archival Information Packages and their supporting Descriptive Information become established within to the ISO 14721 (OAIS) based digital repository.

Internal and External Stakeholders. A digital preservation stakeholder is any organization or individual who can affect or is affected by digital preservation policy, strategy, initiatives, or projects. Broadly speaking, internal references individuals/organizations inside an organization while external references individuals/ organizations outside the organization. ISO 14721 references internal and external stakeholders under the rubric "Defined Community."

Interoperable File Format. *See Open Standard Technology Neutral Format*.

Joint Photographic Experts Group 2000 (JPEG 2000). JPEG 2000 is an international standard (ISO 15444-1) that supports both lossy and lossless compression of digital photographic images. In the context of the Digital Preservation Capability Model it can be a "preferred preservation format."

Legacy Electronic Records. Legacy electronic records are embedded in obsolete software or formats with no backward compatibility or export function to newer software and formats. Legacy records can only be retrieved and rendered by the software application and/or format in which they are embedded or by a viewer. Typically, computer code must be written to transform legacy records into newer, technology neutral open file formats.

Long Term. A period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing user community, on the information being held in a repository. This period extends into the indefinite future.

Long Term Preservation. The combined actions of a preservation repository to ensure that electronic records are accessible, usable, understandable, and trustworthy over technology generations for as long as may be required.

Moving Pictures Expert Group-4 (MPEG-4). MPEG-4 is an International Standard (ISO 14496) for the compression of digital audio content. In the context of the DPCMM it can be considered a "preferred preservation format."

Metadata. Metadata is data (information) about data (information), which is technically correct but simplistic because metadata may serve several purposes in information systems. Descriptive metadata facilitates the search and retrieval of information objects. Administrative metadata supports the management and tracking of information objects. Structural metadata denotes how complex information objects can be reassembled for rendering. Preservation metadata supports activities that ensure the accessibility, usability, understandability, and authenticity of information objects.

Migration. In ISO 14721 migration references actions associated with the transfer of digital content within an ISO 14721 conforming preservation repository. In this context, there are four different actions that may be undertaken. *See Refreshment, Replication, Repackage, and Transformation.*

Motion Joint Photographs Engineering Group (Moving JPEG 2000). Motion JPEG 2000 (MJPEG 2000) is an International Standard (ISO 15444-3) for lossless compression of each video frame in a digital video sequence separately as a JPEG image.

Native File Formats. Native file formats are proprietary formats specific to a software application used to create, store, save, and retrieve electronic records. They are not interoperable in the sense that digital objects embedded in proprietary native file formats can only be "recognized" and opened by the software application used to create and save them unless the software supports an explicit import/export functionality.

Near-preservation ready information. Near preservation ready digital information is encoded in a native, proprietary format but tools exist that can transform it into a technology neutral open standard format. An example is the transformation of Word documents to PDF/A. Some additional processing may be required to assemble the appropriate metadata.

Open Archival Information System (OAIS). An archive, consisting of an organization of people and systems that has accepted the responsibility to preserve information and make it available for a Designated Community. It meets a set of responsibilities, as defined in 3.1 of the ISO 14721:2003 standard that allows an OAIS archive to be distinguished from other uses of the term 'archive'. The term 'Open' in OAIS is used to imply that this Recommendation and future related Recommendations and standards are developed in open forums, and it does not imply that access to the archive is unrestricted.

Open Document Format (ODF). ODF is an International Standards Organization standard, ISO 26300:2000. It is an XML (markup language) standard for creating interoperable office documents, including text, spreadsheets, presentations, and charts.

Open Standard Technology Neutral Format. A technology neutral file format is one that is designed to run on multiple platforms in a variety of software applications. It is an open file format in that the design of the specification involves collaboration in an open, public environment. Open standard technology neutral open file formats can evolve as technology changes and thereby provide a backward compatibility to older versions. Examples of open standards technology neutral file formats are XML and PDF/A.

Plain Text. Plain text is textual material encoded in American Standard Character Interchange (ASCI) without regard for its appearance when rendered. Essentially, plain text is a string of alphanumeric characters with minimal formatting – for example, upper case, lower case, space, spacing, carriage return, $, and * along with alphabetic characters and numbers 0 – 9 among others. Each character is assigned a specific decimal value (A = 65) and a binary value (01000000). Because Plain Text has no formatting functionality like word processing applications, it is interoperable on virtually any technology platform and can be rendered by any text editor.

Portable Network Graphic (PNG). PNG (ISO 15948) is an interoperable international standard lossless compression algorithm for raster images. Among other things, it supports 48 bits of true color and 16 bits per pixel for grayscale raster images.

Preservation Description Information (PDI). A component of Archival Storage in an ISO 14721 conforming repository, PDI contains metadata that is necessary for adequate preservation of the Content Information and which can be categorized as Provenance, Reference, Fixity, and Context information.

Preservation Metadata Implementation Strategy (PREMIS). PREMIS is a standard developed by the Library of Congress that enables designers, managers, and practitioners of digital repositories to have a clear understanding of what a digital preservation system needs in order to execute digital preservation functions. One way this is accomplished is through a Data Dictionary that defines uniform attributes that support an electronic chain of custody that documents the integrity over time as preservation actions are executed.

Preservation Ready Information. Preservation ready information is encoded in a technology neutral open standard format and all necessary metadata has been assembled so that it can be moved (i.e., ingest) into a digital preservation repository with little or no additional processing.

Producer. The role played by those persons, or client systems, who provide the information to be preserved. This can include other OAISs or internal OAIS persons or systems.

Redaction. Historically, redaction describes the process of altering multiple documents slightly and combining them into a single document. Its contemporary meaning refers to concealing content from unauthorized review by obscuring or otherwise deleting specific information that is protected by privacy, proprietary, or national security considerations. Redacted documents are also known as "Disclosure Free."

Refreshment. Refreshment is an ISO 14721 migration activity that references a media instance holding one or more AIPs that is replaced by a media instance of the same type by copying the bits underlying the AIPs. There is no change in Packaging Content, Content Information, Preservation Description Information, and the Archival Storage mapping information or the underlying bit stream of the AIPs information objects

Repackage. Repackage is an ISO 14721 migration activity in an instance in which the replacement of current storage media with different storage media causes in the Content Information and Preservation Description Information. However, there is a change in the bit stream underlying Packaging Information.

Replication. Replication is an ISO 14721 migration activity in which the replacement of storage media of the same type or a new type causes no change in the Content Information and Preservation Description Information. However, there could be a change in the Packaging Content bit streams.

Storage Tier Level. Storage tier levels reference the assignment of different categories of data to different types of storage media in order to reduce total storage cost. Categories may be based on levels of protection needed, performance requirements, frequency of use, and other considerations. Storage tier level considerations will play an increasingly important role when a digital repository has a large volume of digital content (e.g., Terabytes), some of which is accessed frequently and some which is infrequently or not accessed at all.

Submission Agreement: The agreement reached between an OAIS and the Producer that specifies a data model for the Data Submission Session. This data model identifies format/contents and the logical constructs used by the Producer and how they are represented on each media delivery or in a telecommunication session.

Submission Information Package (SIP): An Information Package that is delivered by the Records Producer to the OAIS for use in the construction of one or more AIPs.

Scalable Vector Graphics (SVG). SVG is a World Wide Consortium (W3C) XMD-based markup that supports interoperable two-dimensional vector graphic images. In the context of DPCMM it can be considered a "preferred preservation format."

Surrogate. Surrogate is used in the DPCMM discussion materials to denote a preservation repository that meets some but not all of the specifications of an ISO 14721 conforming repository and to information packages that conform to some but not all of the essential properties of SIPs, AIPs, and DIPs.

Technology Watch Program. Programs such as the United Kingdom National Archives PRONOM program and the Library of Congress Sustainability of Digital Formats Web Site that encompass a variety of tools, and services to support digital preservation functions such as preservation risk assessment, file format sustainability, migration planning, and metadata extraction, among others.

Tagged Image File Format (TIFF). TIFF is an interoperable de facto standard widely used in the capture and storage of digital images. In the context of DPCMM it can be considered a "preferred preservation format."

Transformation. Transformation is an ISO 14721 migration activity associated with replacing current interoperable formats with new interoperable formats to mitigate format obsolescence. There will be changes in the underlying bit streams of Packaging, Information Content, and Preservation Description Information. The resulting AIP is intended to replace the old AIP.

Trustworthy Digital Repository. In ISO 14721 a trusted digital repository is committed to provide long-term access to managed digital resources; accepts responsibility for the long-term maintenance of digital resources on behalf of its depositors and for the benefit of current and future users; designs its system(s) in accordance with commonly accepted conventions and standards to ensure the ongoing management, access, and security of materials deposited within it; establishes methodologies for system evaluation that meet community expectations of trustworthiness; can be depended upon to carry out its long-term responsibilities to depositors and users openly and explicitly; and whose policies, practices, and performance can be audited and measured.

Trustworthy Records. Trustworthy electronic records are reliable and authentic records whose integrity has been preserved over time. Reliability references that records can be trusted as an accurate representation of the activities and facts associated with a transaction(s) because they were captured at or near the time of the transaction. Authenticity means that electronic records are what they purport to be.

Web ARChive (WARC). WARC is an interoperable international standard (ISO 28500) for harvesting, accessing, and exchanging digital content over the Web. In the context of DPCMM it can be considered a "preferred preservation format."

APPENDIX B: Essential Properties of SIPs, AIPs, and DIPs[14]

The 0pen archival information system (OAIS) reference model has three Information Packages Variants:

- Submission Information Packages (SIPs),
- Archival Information Packages (AIPs), and
- Dissemination Information Packages (DIPs).

Electronic records come into an ISO 14721 conforming preservation repository through one or more Submission Information Packages that include the digital content (i.e., the electronic records) and specified administrative, technical, and descriptive metadata. After processing that confirms the accuracy and completeness of the digital content and metadata, the repository accepts the SIPs and transforms them into Archival Information Packages (AIPs).

Initially, an AIP contains all of the SIP content plus any other metadata captured as part of the transfer to Archival Storage. Over time other metadata is added to the AIP that documents any action taken that affects the content of the AIP, including but not limited to device/media renewal and format transformation as new interoperable open standard technology neutral formats displace existing ones.

DIPs, which can be produced on demand or placed on a website for 24/7 access, contain all of the information in selected AIPs plus information that documents when (date and time) the DIP was created or viewed. For confidentiality reasons DIPs contain no information about who may have requested the AIPs.

"Essential properties" should not be confused with functions that generate SIPs, AIPS, and DIPS. Rather, they are evidence or confirmation that these functions have been executed. For example, Confirmation of Integrity (8) means that evidence or documentation of the validation of integrity has been captured and preserved. In this context, the intended purpose of SIP, AIP, and DIP "essential properties" is to facilitate users of the Digital Preservation Capability Maturity Model (DPCMM) in identifying the level of conformance of a preservation repository to SIP, AIP, and DIP specifications and good practices.

These essential properties as evidence of Preservation Repository Actions assume that a PREMIS-like processing plan is in place that defines these preservation actions, identifies the "agent" that conducted them, and when they were conducted. Typically, such plans provide descriptive information, identify restrictions and content access issues, and include processing instructions.

---

[14] We are grateful to Cal Lee (University of North Carolina), Glen McAninch (Kentucky Department for Libraries and Archives), Richard Pearce-Moses (Clayton State University), Chris Prom (University of Illinois), and Matt Veatch (Kansas State Archives) for sharing with us their insights about "essential properties" of SIPs, AIPs, and DIPs.

Essential Property Tables

The next three pages contain tables displaying essential properties of SIPs, AIPs, and DIPs.

Submission Information Packages (SIPs) from Records Producers to an ISO 14721 Preservation Repository

The first column in this table ientifies two broad categories of essential properties – Administration and Content Description. The second column specifies 15 essential properties (metadata) that Record Producers are required to include in a Submission Information Package. The third column identifies evidence that confirms the Preservation Repository has successfully completed appropriate preservation actions.

Note that essential property (16) is the exclusive responsibility of the Preservation Repository.

Archival Information Packages (AIPs) in Archival Storage in an ISO 14721 Preservation Repository

Note that after confirmation of the accuracy and completion of the 16 SIP Essential Properties, the AIP Essential Properties inherit them and Records Producers have no further role in digital preservation. Capturing evidence that confirms the accurate and complete execution of ten specified digital preservation actions adds ten more Essential Properties.

AIPs with these combined 26 Essential Properties are transferred to Archival Storage where they are added to Preservation Description Information (PDI) and bear witness to conformance to ISO 14721 specifications and digital community best practices. Over time other metadata that documents preservation activities (e.g., device/stoarge media renewal) will be added to PDI.

Dissemination Information Packages (DIPs) for Access in an ISO 14721 Preservation Repository

There are no defined specifications for DIPs so these seven Essential Properties of DIPs largely reflect extrapolations from traditional access protocols. As preservation repositories become more robust and users more knowledgeable, it is likely that significant changes will be made to DIP Essential Properties.

Submission Information Packages (SIPs) from Records Producers to an ISO 14721 Preservation Repository

Essential Properties of Information Packages	Records Producer Responsibilities and Metadata	Evidence of Preservation Repository Actions
Administration	(1) Submission Information Agreement including terms binding on both parties	(1) Confirmation of completeness and accuracy of Submission Information Agreement terms
	(2) Transfer of legal custody to the archives, including date and authority	(2) Confirmation of Acceptance of Legal Transfer
	Identification of Intellectual Property Rights (3) Copyright (4) Access Restrictions	(3) Confirmation of accuracy and completeness of Rights (4) Confirmation of accuracy and completeness of Access Restrictions
	(5) Identification of Volume (Bytes)	(5) Confirmation of accuracy and completeness of volume
	(6) Identification of Record Type(s)	(6) Confirmation of accuracy and completeness of Record Type(s)
	(7) Identification of Format	(7) Confirmation of accuracy and completeness of Format identification
	(8) Identification of Integrity (hash digest)	(8) Confirmation of Integrity (hash digest)
	(9) Virus free SIP	(9) Confirmation of Virus quarantine results
Content Description	(10) Identification of Producer (e.g., Contributor)	(10) Confirmation of accuracy and completeness of identity of Producer (e.g., Contributor)
	(11) Identification of Title	(11) Confirmation of accuracy and completeness of Title
	(12) Identification of Date of creation	(12) Confirmation of Date of creation
	(13) Identification of Date of Transfer	(13) Confirmation of Date of transfer
	(14) Identification of Provenance (e.g., relations, arrangement)	(14) Confirmation of accuracy and completeness of provenance
	(15) Identification of Software, including version, used to create or capture digital objects	(15) Confirmation of accuracy and completeness of identification of software, including version, used to create or capture digital objects
	NA	(16) Confirmation of Assignment of Unique Persistent Identifier

Archival Information Packages (AIPs) in Archival Storage in an ISO 14721 Preservation Repository

Essential Properties of Information Packages	Records Producer Responsibilities and Metadata	Evidence of Preservation Repository Actions
SIP Administration	NA	Include All Previous SIP Administration Essential Properties
SIP Content Description	NA	Include All Previous SIP Content Description Essential Properties
AIP Creation Preservation Actions	NA	(17) Confirmation of accuracy completeness of Transformation to "Preferred File Format"
	NA	(18) Confirmation of accuracy and completeness of Transformation into AIPs
	NA	(19) Confirmation of accuracy and completeness of Transfer to Archival Storage
	NA	(20) Confirmation of Identity of Agent Responsible for Device/Media Renewal and Date of Action
	NA	(21) Confirmation of accuracy and completeness of Technical Specifications for Current and New Device/Media, Software and Operating System Used
	NA	(22) Confirmation of Integrity of scheduled Device/Media Renewal with Pre- and Post-Action Cryptographic Hash Algorithm
	NA	(23) Confirmation of Identity of Agent Responsible for Format Transformations and Date of Action
	NA	(24) Confirmation of technical information about Format Transformations, Including Software and Operating System Used
	NA	(25) Confirmation of Integrity of Format Transformations with Super Hash or Encapsulated Digital Signature and Date of Action
	NA	(26) Confirmation of accuracy and completeness of documentation of Preservation Actions in Preservation Description Information (PDI)
	NA	(27) Confirmation of accuracy and completeness of Error Logs where appropriate for all AIPs and Digital Preservation Actions

Dissemination Information Packages (DIPs) for Access in an ISO 14721 Preservation Repository

Essential Properties of Information Packages	Records Producer Responsibilities and Metadata	Evidence of Preservation Repository Actions
DIP Access Actions	NA	(28) Confirmation of User request for specific AIPs
	NA	(29) Confirmation that no restrictions limit access to selected AIP(s)
	NA	(30) Confirmation that if necessary Redaction of AIPs to Protect Confidentiality was correctly executed
	NA	(31) Confirmation of any changes made in one or more AIPs to facilitate access
	NA	(32) Confirmation that a request was sent to Archival Storage to Produce one or more DIPs
	NA	(33) Confirmation of actions taken to produce DIPs
	NA	(34) Confirmation of preservation of date and time of delivery of DIPs

###