GUIDANCE

2022 CoSA Digital Preservation Capability

Self-Assessment Survey Glossary

*Based on the Digital Preservation Capability Maturity Model (DPCMM) by Charles M. Dollar and Lori J. Ashley and customized for the benefit of the members of the Council State Archivists (CoSA)*

**Access**

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

**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.

**Conforming AIPs**

Conforms to the ISO 14721 (OAIS) specifications for Archival Information Packages (AIPs).

**Conforming SIPs**

Conforms to the ISO 14721 (OAIS) specifications for Submission Information Packages (SIPs).

**Consumer**

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

**Designated Community**

An identified group of one or more records Producers who create electronic records of long-term value that will be transferred to the trust repository along with potential Consumers who should be able to understand a particular set of information. The potential Consumers 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**

Disclosure free references the practice of allowing the disclosure of information in a document while blocking the disclosure of personally identifiable information or sensitive confidential information in the same document.

**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**

A U.S. Department of Defense (DoD) Directive that identifies the functional requirements and specifications for electronic records management software applications.

**Electronic Record**

Digital information content that is captured and stored in a computer storage device or media for future use as evidence of business transactions that requires access to computer technology to render it intelligible to humans. "Born digital" refers to records created in a digital format while scanned digital records are reproductions or images of hard copy records.

**Encapsulation**

For digital preservation purposes encapsulation involves the use of an electronic wrapper or container in which records, and all of their associated metadata are placed for storage. An AIP represents one form of encapsulation. Another form of encapsulation involves creating an XML based "wrapper" that is self- extracting and then affixing a digital signature to the wrapper that supports authentication of a preservation activity.

**Fixity**

Fixity information provides documentation that no unauthorized changes have occurred to the contents of AIPs during the execution of preservation activities.

**Hash Digest Algorithm (Cryptographic)**

A cryptographic hash algorithm takes any digital object regardless of size or content type and normalizes it to a fixed length bit stream (e.g., 128 bits). This fixed length bit stream is called a hash digest and it serves as a "digital fingerprint" of a larger digital object. Cryptographic hash algorithms are used to detect accidental or intention corruption in digital objects and to authenticate digital signatures: the change of a single bit in the original digital object will result in a different hash digest. It is computationally infeasible for two different digital objects to have the same hash digest or to reconstruct a data object from this hash digest.

**Information Package (ISO 14721)**

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.

**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**

Plans for different types of records, models for preservation approaches and criteria, and a framework of repository components and services will require tighter cooperation and engagement between long- standing partners such as IT, peer archives/RM units, software and service providers, and other support functions. Interdependencies between and among the operations of records producing units of government, legal and statutory requirements, information technology policies and governance, and historical accountability should be systematically addressed.

###### ISO 14721

ISO 14721 is an international standard for a reference model of an open archival information system (OAIS). It establishes a framework for an organizational scheme composed of people who accept the responsibility to preserve information and make it available to a designated community. The reference model addresses a full range of archival information preservation functions including ingest, archival storage, data management, access, and dissemination. It also addresses the migration of digital information to new media and formats.

###### ISO 16363

ISO 16363 is an international standard for audit and certification of trusted digital repositories. It establishes a framework that includes performance metrics that are empirically derived for validating that a digital repository conforms to the specifications.

**ISO 14721/ISO 16363 Conforming Archival Repository**

An archival repository that conforms with the specifications of ISO 14721 and ISO 16363.

**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 electronic 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 digital information 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 act of maintaining trustworthy digital information in a correct and independently understandable form for as long as required

**MD5 Hash Digest**

MD5 is a cryptographic algorithm that compresses any digital object to a 128-bit value. A change in only one bit of a digital object will result in different hash value. It is called cryptographic because it was designed to make it computationally infeasible for two different digital objects to have the same hash digest value. However, researchers have demonstrated how it is possible to create two digital objects that share the same MD5 hash value. The U.S. Government advises against the use of MD5 for sensitive information, including digital signatures.

**Native File Formats**

Electronic records in a native file format can only be recognized and opened by the software application that originally was used to create the records. Sometimes an application other than the original software application may be able to open records in a native format but key features (line spacing, special type fonts, and the like) may be rendered differently.

**Near Preservation-Ready Born Digital 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 Standard Technology Neutral File (OS/TN) 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. Technology neutral open file formats can evolve as technology changes and thereby provide a backward compatibility to older versions. Examples of technology neutral file formats are XML and PDF/A.

**Partially Conforming**

Partially conforming approaches and techniques to addressing digital preservation requirements are substantive and represent evolving/emerging capabilities. They do not, however, fully meet ISO 14721/ISO 16363 specifications.

**PREMIS-based Data Dictionary**

PREMIS (Preservation Metadata Information Strategy) 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 to know 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 Description Information (PDI)**

The information which is necessary for adequate preservation of the Content Information and which can be categorized as Provenance, Reference, Fixity, and Context information.

**Preservation Ready Born Digital Information**

Preservation ready born digital 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 that provide the information to be preserved. This can include other archival information systems or internal persons or systems.

**Raster Bit Map Image**

A raster image, which is also called a bitmap, is a representation of an image as rectangular matrix that contains a specified sequence of rows of small bits of information (pixels or dots). Usually, this sequence is defined as a number of dots per inch and whether the image is bi-tonal gray, or color. A 300 dots per inch resolution of a black/white image means that there are 300 dots per inch horizontally and vertically in each square inch of the image. A 600 dots per inch color digital photograph supports up to 24 color values for each bit in a square inch of an image. Unlike a vector graphic image, a raster image consists solely of a sequence of pixels that must be rendered into human understandable form by a display device or printer. The raster image formats widely used today include Joint Photographic Expert Group (JPEG 2000), Tagged Image File Format (TIFF), and Portable Network Graphics (PNG).

**Standard Hash Algorithm 1 (SHA-1)**

A cryptographic hash digest developed by the U.S. National Security Agency A 160-bit hash function to be part of the Digital Signature Algorithm. The hash digest contains 160 bits, which increases the computational infeasibility of two different digital objects having the same hash value. However, researchers have demonstrated that it in rare circumstances it is computationally feasible for two different digital objects to have the same SHA-1 hash digest.

**Standard Hash Algorithm 2 (SHA-2)**

Designed by the National Security Agency, it has a 256-bit hash value, which makes it the most powerful hash digest algorithm currently available.

**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 Producer to the OAIS for use in the construction of one or more AIPs.

**Trustworthy Digital Repository**

A trustworthy digital repository accepts responsibility for the long-term maintenance of digital records for current and future users; has an organizational system that supports the long-term sustainability of the repository and its contents; designs and implements its systems in such a way as to ensure on-going access to and security of digital records in its custody; establishes credible methodologies for system evaluation that meet community expectations of trustworthiness; and supports policies, practices, and actions that can be measured and audited.

**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.

**Vector Graphics**

Vector graphics images consist of lines, colors, curves or any other geometrical shapes and attributes that are stored as mathematical expressions, such as where a line begins, its shape, where it ends, and its color. Changes in the mathematical expressions will result in changes in the image. Unlike raster images, there is no loss of clarity of a vector graphics image when it is made larger.

**Viewer Technology**

A software application that enables viewing (render on monitor, print, or copy) digital records in their original format without having access to the application used to create the digital records. Outside In (Oracle) viewer technology supports rendering capability of more than 500 file formats. Reliance on viewer technologies to render digital records in native and legacy file formats is a temporary solution to file format obsolescence.