• OPEN REPOSITORIES REPORT SERIES

Digital Preservation Functionality in Canadian Repositories

by *Tomasz Neugebauer, Pierre Lasou, Andrea Kosavic, and Tim Walsh*

(on behalf of the CARL Open Repositories Working Group's Task Group on Next Generation Repositories)

DECEMBER 2019

www.carl-abrc.ca

CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES • ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA

Table of Contents

Introduction 2

Summary of Recommendations 3

Digital Preservation Requirements 3

Bit preservation 4

Preservation Metadata 4

AIP Export 5

Open Repository Systems in Canada 5

DSpace 6

EPrints 6

Islandora 7

Samvera 8

Omeka 9

ContentDM 10

DigitalCommons 10

Glossary of Digital Preservation Terms 12

Appendix 1. Summary of Digital Preservation Functionality in Repository Systems . 14

Introduction

Digital preservation is the series of management policies and activities necessary to ensure the long-term usability, authenticity, discoverability and access of digital content. Digital preservation is necessary to deal with challenges such as file corruption, media failure, and technological change that may affect the continued usability of digital content over time. Digital Preservation functionality is listed as one of the recommendations in the Behaviours and Technical Recommendations of the COAR Next Generation Repositories Working Group[1].

In May 2018, The Portage Network released a white paper entitled Research Data Preservation in Canada[2]. The Portage white paper offers an overview of digital preservation goals, principles and organizations working towards building a sustainable digital preservation infrastructure in Canada. It proposes a networked model where repositories function as nodes responsible for ingest, access and data management, augmented with the development of preservation service nodes that offer archival storage.

This Digital Preservation Functionality in Canadian Repositories report was prepared as a part of the work of CARL Open Repositories Working Group, Task Force on Next Generation Repositories. The report was written to make progress on one of the key next steps in achieving the vision outlined in the Portage white paper: the building of a common understanding of basic digital preservation requirements and functionality necessary to achieve this vision of a sustainable digital preservation network. This report focuses on technical functional requirements, and as such, it is intended for scholarly communication librarians and repository administrators assessing or looking to enhance the digital preservation support in their repositories. We begin by listing the digital preservation requirements that must minimally exist in repositories as places of ingest and management of content. We then proceed to describe these in more detail, providing a use case for each, and a summary of the support for these in each of the repository systems that have been identified as in common use in Canada. This report focuses on the technical requirements, but it is important to emphasize that the creation and maintenance of policies and activities for digital preservation also requires that organizations that maintain repositories define and establish roles and responsibilities for this purpose.

[1] Rodrigues, E., & Shearer, K. (2017). Next generation repositories: Behaviours and technical recommendations of the COAR Next Generation Repositories Working Group. https://www.coar-repositories.org/files/NGR-Final-Formatted-Report-cc.pdf

[2] Qasim, U., Davis, C., Garnett, A., Marks, S., & Moosberger, M. (2018). Research data preservation in Canada: A white paper. https://portagenetwork.ca/wp-content/uploads/2018/05/Portage-PEG-WhitePaper-EN.pdf

Summary of Recommendations

In this report, and in light of the vision of a networked digital preservation system in Canada, we attempt to establish minimal recommended level of digital preservation functionality in repository systems (Digital Preservation Requirements). Building digital preservation capacity requires organizational commitment through the recognition of new roles and the development of digital preservation policies and technologies. Focusing specifically on the technological aspects of digital preservation in repository systems, we recommend that repository librarians, managers and administrators continue to build an understanding of which digital preservation processes and functionalities are currently offered in their systems. This understanding of the digital preservation functionality in repository software systems is useful in identifying and addressing achievable areas of improvement, such as: configuring existing digital preservation plugins, services or command line scripts, promoting the development of additional functionalities that are not currently available, ensuring that storage and backup procedures meet minimal digital preservation requirements, and developing a digital preservation policy and organizational awareness and commitment.

Digital Preservation Requirements

As the point of ingest and site of data management for digital objects of significant cultural value, repositories need to have some digital preservation functionality built in and should interoperate with current and future dedicated digital preservation services. The functions and responsibilities involved in digital preservation are well-described within standards such as the National Digital Stewardship Alliance (NDSA) Levels of Digital Preservation[3] and the Open Archival Information System reference model (ISO 14721)[4].

Within the context of repository systems which are likely to work with external digital preservation services, minimal digital preservation capacity involves ensuring that:

- files are not lost or corrupted (*bit preservation*)
- sufficient information about the technical characteristics, chain of custody, and provenance of digital objects is collected to enable future preservation actions and access (*preservation metadata*)
- objects and metadata can be passed to dedicated digital preservation systems when appropriate (*AIP export*)

[3] National Digital Stewardship Alliance – Digital Library Federation. (2019). Levels of digital preservation v2.0. https://ndsa.org/activities/levels-of-digital-preservation/

[4] OAIS (2012) CCSDS 650.0-M-2: Reference model for an open archival information system (OAIS). Magenta Book. Issue 1. June 2012 (ISO 14721:2003) http://public.ccsds.org/publications/archive/650x0m2.pdf

Repository managers may also want to ensure that files can continue to be rendered and meaningfully accessed over time despite the effects of file format, software, and hardware obsolescence. It may be appropriate to address solutions to this issue, including *file format migration*, *normalization*, and *emulation* within repositories or by exporting content to digital preservation service providers that have the capacity to enact preservation planning and such strategies at scale.

Bit preservation

Use Case: *A digital object uploaded to a repository becomes corrupted during an upgrade of the disks that the repository is residing on. The checksum of the object was stored on ingest, and a periodic fixity check identified the problem with the object to the repository administrator. The administrator is then able to retrieve the uncorrupted file from one of the backup copies before the backup is overwritten*.

Bit preservation is concerned with ensuring that digital objects remain demonstrably authentic over time. The long-term threats to digital objects are well documented and include media failure, hardware failure, software failure, natural disaster, external and internal attacks, and economic or organizational failure.[5]

Technical approaches to bit preservation mitigate these risks by storing multiple independent copies of data, using resilient storage technologies that take advantage of techniques such as erasure coding to detect and repair changes to objects at a bit- or block-level, and running fixity checks (periodically verifying that checksums for digital objects in storage match previously recorded values) to verify that files have not corrupted or otherwise changed. Fixity checking events should be stored in logs and recorded in preservation metadata for the digital objects. Approaches to storage must balance resiliency with organizational and budgetary sustainability to minimize the risk of digital preservation activities becoming prohibitively expensive or otherwise organizationally infeasible.

Minimally, digital objects of any significant value should always be stored and backed up following the 3-2-1 rule (three copies of data, on two types of storage media, with at least one off-site backup). The NDSA Levels of Digital Preservation offers an approachable metric for how to incrementally improve storage from a digital preservation perspective[6].

Preservation Metadata

Use Case: *A thesis depositor is required to upload an archival PDF/A, but uploads a PDF document with multiple formatting problems instead, including the use of external content sources and some custom fonts only available on the depositor's computer. The PDF document renders on the depositor's computer, but it includes text that is unreadable when opened on other devices. The repository attempts to*

[5] Rosenthali, D. S. H., Robertsoni, T., Lipkisii T., Reichi V., Morabitoi, S. (2005). Requirements for digital preservation systems : A bottom-up approach. *D-Lib Magazine, 11*(11). http://www.dlib.org/dlib/november05/rosenthal/11rosenthal.html

[6] National Digital Stewardship Alliance – Digital Library Federation. (2019). Levels of digital preservation v2.0. https://ndsa.org/activities/levels-of-digital-preservation/

validate the PDF/A file format of the object on ingest and reports to the depositor or repository editor that the uploaded object fails to conform to the required format for theses deposits. The depositor is then required to correct the formatting issues before the thesis object is accepted.

Typically expressed using the PREMIS (Preservation Metadata: Implementation Strategies) Data Dictionary and often within the context of a METS file, preservation metadata "documents the technical processes associated with preservation, specifies rights management information, establishes the authenticity of digital content, and records the chain of custody and provenance for a digital object."[7]

In addition to implementing PREMIS, repository systems can support preservation metadata by robustly characterizing digital objects. This characterization includes identifying and validating file formats and by extracting features (technical metadata about files, e.g. codecs and bit rates) from digital objects.

AIP Export

Use Case: *The current repository software platform does not offer a satisfactory level of digital preservation functionality, and a new consortial digital preservation service node becomes available. The AIP export functionality allows the repository manager to export all of the objects and metadata from the repository to the new digital preservation service. If necessary, the AIP export also allows the repository manager to migrate all of the contents to a new access platform that includes the minimal requirements for preservation*.

AIP export functionality is required for interoperability with dedicated digital preservation systems. It may be possible to extend the core functionality of repository systems to include the full suite of digital preservation support, but digital preservation is an ongoing process that requires ongoing evaluation of risks that include the platform it resides on. Therefore, the functionality of exporting the digital objects along with metadata is required in repository systems. This functionality should take advantage of standards such as METS and PREMIS whenever possible.

Open Repository Systems in Canada

Based on the ORWG's 2018 survey of Canadian repositories, *Repository landscape in Canada*[8], the following is a list of systems that Canadian open repositories are primarily running on: DSpace, EPrints, Islandora, Samvera, Omeka, ContentDM, DigitalCommons. The following table contains a summary of digital preservation

[7] National Digital Stewardship Alliance - Digital Library Federation. (n.d.) Glossary. https://ndsa.org/glossary/

[8] *This dataset provides an overview of the repository systems used in Canada (repository name, institution, URL, OAI server, software and versions). Information was gathered in October 2018.* https://drive.google.com/open?id=1WjY9mRL6HNQ1hOf2LeedK33J8npdjLneCkWp-raOrKQ

functionality in each system, followed by a short description of digital preservation support in each system.

DSpace

Preservation Metadata

Built in features for robust file characterization are not included in DSpace. Bitstream Format Profiler[9], a part of the Curation Tasks module, performs format identification by file extension only (not format signature) for digital objects on ingest.

"DSpace has defined a Technical Metadata Element Set to fulfill their preservation and content lifecycle management information needs. This Technical Metadata Element Set is best expressed using the PREMIS Preservation Metadata Schema: Object. DSpace use of the PREMIS Data Dictionary to represent needed technical metadata elements does not constitute full implementation of the PREMIS data model."[10]

Bit Preservation

DSpace generates an MD5 checksum on ingest for each digital objects. Since version 1.7, DSpace includes a Checksum Checker[11] program can be run from the command line and set to run periodically as a cron task to verify the checksums for all stored objects.

It is up to DSpace repository administrators to ensure that storage and backups meet minimal requirements for digital preservation.

AIP Export

Since version 1.7, DSpace has AIP export[12] functionality built in. This export format may be used for import in dedicated digital preservation systems (i.e. Archivematica).

EPrints

Most of the digital preservation functionality available to the EPrints community is the result of the 2005-2009 Preserv project [13]. The Digital Preservation Toolkit[14] that was

[9] DSpace 6.x documentation, curation system: Bitstream format profiler. https://wiki.duraspace.org/display/DSDOC6x/Curation+System#CurationSystem-BitstreamFormatProfiler

[10] DSpace development projects, DSpaceMETSSIPProfile. https://wiki.duraspace.org/display/DSPACE/DSpaceMETSSIPProfile

[11] DSpace 5.x documentation, system administration, validating checksums of bitstreams. https://wiki.duraspace.org/display/DSDOC5x/Validating+CheckSums+of+Bitstreams

[12] DSpace 6.x documentation, DSpace AIP format. https://wiki.duraspace.org/dsdoc6x/system-administration/aip-backup-and-restore/dspace-aip-format#DSpaceAIPFormat-MakeupandDefinitionofAIPs

[13] Preserv: Repository preservation and interoperability. http://preserv.eprints.org/

[14] Tarrant, D. (2012). Preservation toolkit. http://bazaar.eprints.org/142/

released in 2011 is out-of-date and incompatible with the latest version of dependencies (e.g. DROID).

Preservation Metadata

EPrints performs file format identification for all files using the Unix built-in File utility and performs some feature extraction from audiovisual formats using ffprobe/ffmpeg.

EPrints does include a METS export plugin, but the METS files contain insufficient digital preservation metadata. It also includes a "history" module that tracks changes to metadata for each item, however, it does not implement PREMIS.

Bit Preservation

EPrints generates a checksum for files on ingest, but tests reveal that not all objects are assigned a checksum. There is no periodic fixity checking built into EPrints.

It is up to EPrints repository administrators to ensure that storage and backups meet minimal requirements for digital preservation.

AIP Export

The EPrints community is working on an export plugin to leverage the digital preservation Archivematica software. The proposed integration is to export out of EPrints all live archive eprints and package them along with the digital files for preservation in Archivematica[15].

Islandora

Currently, the Islandora 7 stack includes Fedora 3, but the next generation version of Islandora currently in development (Islandora 8) will replace this with Fedora 4 and 5. Much of Islandora's digital preservation capabilities are integrated through optional Utility Modules and Solution Packs.[16] It is up to repository managers to integrate, enable, and configure these modules to support digital preservation processes in Islandora.

Preservation Metadata

With the Islandora FITS Utility Module, Islandora will robustly characterize and extract features from ingested digital objects using the open source File Information Tool Set (FITS).[17]

Bit Preservation

Islandora supports checksum generation and ongoing fixity checking via the Islandora Checksum and Islandora Checksum Checker Utility Modules. The results of fixity checks over time are recorded in PREMIS using the Islandora PREMIS Utility Module.

[15] Digital preservation through EPrints-Archivematica integration - an EPrints export plugin to Archivematica. https://github.com/eprintsug/EPrintsArchivematica

[16] Islandora documentation. https://wiki.duraspace.org/display/ISLANDORA/

[17] File Information Tool Set (FITS). https://projects.iq.harvard.edu/fits/home

It is up to Islandora repository administrators to ensure that storage and backups meet minimal requirements for digital preservation.

AIP Export

Objects can be exported from Islandora in the preservation-friendly BagIt packaging format using the Islandora BagIt Utility Module.

If a site also implements Archivematica, it's currently possible to push Islandora objects to Archivematica via the Archidora module.[18] In this workflow, the Islandora objects become a SIP which is ingested into Archivematica and subsequently turned into an AIP and DIP according to configuration in Archivematica.

Samvera

Samvera is an open source repository framework built on Fedora 4, Ruby on Rails, Solr, and Blacklight. The Samvera framework "consists of a number of Ruby gems that can be combined, configured and adapted to serve a wide variety of needs".[19] A popular way of implementing Samvera is through "solution bundles" such as Avalon Media System, Hyrax, and Hyku (the product resulting from the "Hydra-In-A-Box" project). As such, repositories built on the Samvera framework may differ significantly in terms of supported digital preservation functionality. However, all Samvera repositories leverage Fedora's digital preservation features, including integrated checksums, item-level version control with rollback, and a complete audit history.

Preservation Metadata

The hydra-file_characterization module[20], a Core Component of the Samvera community, characterizes ingested digital objects using FITS. Hyrax includes this component by default.

Some Object metadata (e.g. checksum value/message digest) is expressed using elements from the PREMIS data dictionary, but Samvera repositories do not include a full PREMIS implementation.

Bit Preservation

Fixity checking (re-calculating and comparing checksums) is available via the Fedora API. However, this is not accessible through the Hyrax web admin interface. It would have to be run from the command line by a developer or sysadmin. Further information is available in the Fedora 5 documentation.[21]

[18] Islandora documentation: Archidora. https://wiki.duraspace.org/display/ISLANDORA/Archidora

[19] Samvera. https://samvera.org/samvera-open-source-repository-framework/

[20] Samvera file characterization (extracted from Sufia). https://github.com/samvera/hydra-file_characterization

[21] Fedora 5.x documentation: Fixity checking. https://wiki.duraspace.org/display/FEDORA5x/Fixity+Checking

As of 2018, Hyrax does not include an admin UI for auditing Fedora's built-in fixity checks, or for restoring objects from backups if bitrot or file corruption is detected.[22]

AIP Export

University of York (UK) and The University of Hull are working on Samvera/Archivematica integration, but no open plugins are available.

It is up to Samvera repository administrators to ensure that storage and backups meet minimal requirements for digital preservation.

Omeka

As a platform for online presentation and exhibitions, Omeka does not offer native support for full-lifecycle digital asset management and digital preservation. Administrators of Omeka systems should ensure that digital content requiring preservation is managed using separate systems and processes.

Preservation Metadata

The Omeka source code includes the Zend Framework, which contains PHP libraries for identifying and storing MIME types and checksums (using the CRC32, MD5, and SHA-1 algorithms) and performing some basic file characterization. Omeka source code also includes getID3, a PHP script that performs some file characterization of multimedia file formats. It is unclear to what degree these utilities are used to routinely characterize files uploaded to Omeka.

Omeka does not implement PREMIS.

Bit Preservation

Omeka appears to record checksums for uploaded files, but does not contain functionality to conduct routine fixity checks over time.

It is up to Omeka repository administrators to ensure that storage and backups meet minimal requirements for digital preservation.

AIP Export

Integrations with other repository systems have focused on importing content to Omeka for presentation. There are import modules for Fedora and DSpace. Export functionality is much more limited.

[22] Rochkind, J. (2017). Exploring and planning with Sufia/Hyrax/Fedora fixity validation. https://bibwild.wordpress.com/2017/05/01/exploring-and-planning-with-sufiahyraxfedora-fixity-validation/

ContentDM

As ContentDM is proprietary software by OCLC, information about digital preservation functionality is drawn from marketing materials[23] and cannot be verified from the source code.

Preservation Metadata

ContentDM performs virus checking as well as some degree of file format identification (including verification that the file format matches the file's extension) for uploaded files.

ContentDM does not implement PREMIS.

Bit Preservation

ContentDM generates and monitors checksums for uploaded files and makes "health reports" available to administrators which include a periodic fixity check.

Files uploaded to ContentDM are stored in ISO 27001-certified data centres. Storage details such as the number and geographic locations of copies are unknown.

AIP Export

Export functionality is limited to descriptive Dublin Core metadata into various formats such as XML and tab-delimited, but object-level preservation metadata such as checksums do not appear to be included. One institution describes[24] requesting from OCLC Support and downloading a tar export file with the objects.

DigitalCommons

As DigitalCommons[25] is proprietary software by Bepress (RELX), information about digital preservation functionality is drawn from marketing materials and cannot be verified from the source code. The marketing materials also clarify that some of the digital preservation functionality is actually sold as a separate bepress Archive[26] service made available in 2016.[27]

[23] CONTENTdm. Digital preservation: Keep your digital special collections secure. OCLC. https://www.oclc.org/content/dam/oclc/services/brochures/215630-WWAE_Cdm-DigPres-Feature-Flier.pdf

[24] Bullen, A. (2016). A Doomsday scenario: Exporting CONTENTdm records to XTF. *D-Lib Magazine, 22*(11). http://www.dlib.org/dlib/november16/bullen/11bullen.html

[25] Bepress Digital Commons. https://www.bepress.com/products/digital-commons/

[26] Getting started with Bepress Archive. https://www.bepress.com/reference_guide_dc/getting-started-bepress-archive/

[27] Palmer, L. A. (2017). Storage made simple: Preserving digital objects with bepress Archive and Amazon S3. Northeast Institutional Repository Day. https://doi.org/10.13028/trd9-mr81. Retrieved from https://escholarship.umassmed.edu/neirug/2017/program/11.

Preservation Metadata

DigitalCommons does not implement PREMIS. DigitalCommons stores the mime type of uploaded files. The software is not open source, so it is not possible to verify the tool used for the mime type identification.

DigitalCommons stores objects in their original format. The "option to build in capabilities" for emulation and migration are mentioned, but the current stage of development for these functionalities in DigitalCommons is unknown. Bepress also claims to be "committed to making PDFs web-accessible on a permanent basis", promising to migrate PDF objects in case "Adobe changes Acrobat to such an extent that older PDFs are no longer readable."[28].

Bit Preservation

"Production files" uploaded to DigitalCommons are stored on redundant storage, and backed up through "hourly and daily snapshots" and periodic offsite backups. A second copy of the offsite backup is sent to Amazon S3 cloud storage. Another backup is sent to Amazon Glacier, which "performs regular, systematic data integrity checks and is built to be automatically self-healing."[29] However, the subscriber can only have access to a preservation backup if they also subscribe to the bepress Archive service.

Bepress claims to be "LOCKSS-compliant"[30], encouraging customers to join a network for preservation.

AIP Export

Bepress refers to OAI-PMH as the method to transfer scholarship off the platform [31] while one DigitalCommons subscriber developed software to address the "difficulty of batch exporting object files."[32]

The bepress Archive service in combination with an Amazon S3 subscription provides institutional access to Dublin Core metadata along with the most recent versions of uploaded digital objects in a file/folder hierarchy. This service includes checksums and

[28] Format migration and emulation. Safeguarding your content with Digital Commons. https://www.bepress.com/reference_guide_dc/safeguarding-content-digital-commons/#format-migration-and-emulation

[29] A Robust infrastructure and dedicated staff to manage it. Safeguarding your content with Digital Commons. https://www.bepress.com/reference_guide_dc/safeguarding-content-digital-commons/#a-robust-infrastructure-and-dedicated-staff-to-manage-it

[30] Safeguarding your content with Digital Commons. https://www.bepress.com/reference_guide_dc/safeguarding-content-digital-commons/

[31] Bepress hosted & cloud-based. https://www.bepress.com/hosted-cloud-based/

[32] Migration guide: Digital Commons. https://github.com/MarcusBarnes/mik/wiki/Migration-Guide:-Digital-Commons

fixity checking[33], but it is unclear if the checksum metadata can be exported. A METS file is not available.

Glossary of Digital Preservation Terms

- Characterization: “Characterisation is comprised of four elements: identifying the object's format; validating that the object conforms to its format's technical norms; extracting technical metadata from the object; and assessing whether the object should be accepted into a repository, based on policies set by the curator”[34]
- Identification: “Identification is the process of determining the *presumptive* format of a digital object on the basis of suggestive extrinsic hints and intrinsic signatures, both internal (e.g. magic number) and external (e.g. file extension).”[35]
- Validation: “Validation is the process of determining the level of *conformance* of a digital object to the normative syntactic and semantic rules defined by the authoritative specification of the object's format.”[36]
- Feature extraction: “Feature extraction is the process of reporting the *intrinsic properties* of a digital object significant to preservation planning and action.”[37]
- Checksum: “An algorithmically-computed numeric value for a file or a set of files used to validate the state and content of the file for the purpose of detecting accidental errors that may have been introduced during its transmission or storage. The integrity of the data can be checked at any later time by recomputing the checksum and comparing it with the stored one.”[38]
- Fixity checking: “A mechanism to verify that a digital object has not been altered in an undocumented manner. Checksums, message digests and digital signatures are examples of tools to run fixity checks. Fixity information, the information created by these fixity checks, provides evidence for the integrity and authenticity of the digital objects and are essential to enabling trust.”[39]

[33] Safeguarding your content with Digital Commons. https://www.bepress.com/reference_guide_dc/safeguarding-content-digital-commons/

[34] Digital Preservation Coalition. Glossary. Digital preservation handbook. https://www.dpconline.org/handbook/glossary

[35] JHOVE2 documentation. https://bitbucket.org/jhove2/main/wiki/JHOVE2_Frequently_Asked_Questions_(FAQ)

[36] Ibid

[37] Ibid

[38] National Digital Stewardship Alliance - Digital Library Federation. Glossary. https://ndsa.org/glossary/

[39] Ibid

- AIP: "*Archival Information Package (AIP)*: An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an OAIS [Open Archival Information System]."[40]
- METS: "The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library."[41]
- **Preservation metadata**: "The contextual information necessary to carry out, document, and evaluate the processes that support the long-term retention and accessibility of digital content. Preservation metadata documents the technical processes associated with preservation, specifies rights management information, establishes the authenticity of digital content, and records the chain of custody and provenance for a digital object."[42]
- PREMIS: "The PREMIS Data Dictionary for Preservation Metadata is the international standard for metadata to support the preservation of digital objects and ensure their long-term usability. Developed by an international team of experts, PREMIS is implemented in digital preservation projects around the world, and support for PREMIS is incorporated into a number of commercial and open-source digital preservation tools and systems."[43]
- **Archival storage**: "The category of digital storage that provides the services and functions for the long-term storage, maintenance and retrieval of digital objects."[44] Archival storage requires multiple non-collocated copies of stored data and should consider risks such as media failure, natural disaster, obsolescence, and vendor lock-in. For further detail, see the NDSA Levels of Digital Preservation.[45]

[40] OAIS (2012) CCSDS 650.0-M-2: Reference model for an open archival information system (OAIS). Magenta Book. Issue 1. June 2012 (ISO 14721:2003) http://public.ccsds.org/publications/archive/650x0m2.pdf

[41] Metadata Encoding and Transmission Standard (METS). Library of Congress. http://www.loc.gov/standards/mets/

[42] National Digital Stewardship Alliance - Digital Library Federation. Glossary. https://ndsa.org/glossary/

[43] PREMIS Data Dictionary for Preservation Metadata (PREMIS). Library of Congress. https://www.loc.gov/standards/premis/

[44] National Digital Stewardship Alliance - Digital Library Federation. Glossary. https://ndsa.org/glossary/

[45] National Digital Stewardship Alliance – Digital Library Federation. (2019). Levels of digital preservation v2.0. https://ndsa.org/activities/levels-of-digital-preservation/

Appendix 1. Summary of Digital Preservation Functionality in Repository Systems

Repository Software	Preservation metadata				Bit preservation		AIP export (objects + METS)
	Identification	Validation	Feature extraction	PREMIS	Check-sum on Ingest	Fixity Checking	
DSpace	yes (file extension only)	no	no	Partial. Uses some PREMIS Object elements but no Event implement-ation.	yes	Via command line script	yes
EPrints	yes (File utility)	no	yes (some formats)	no	yes	no	Partial (METS export but no AIP export)
Islandora *via Utility Modules ** via Archidora	yes*	?	yes*	yes*	yes*	yes*	yes**
Samvera	yes	?	yes (some formats)	Partial. Uses some PREMIS Object elements but no Event implement-ation.	yes	Via command line script	no?
Omeka	yes	By file extension and mime type only	Yes (some formats)	no	yes	no	no
ContentDM	yes	?	?	no	yes	yes (details unknown)	no

DigitalCommons * (in Amazon Glacier and S3) but client only has access to a preservation copy if they purchase an additional Archive service	yes	?	?	no	yes *	yes *	no