Appendix A: Glossary

Access model: Combination of environment and processes that enables users to access content and the archival institution to control and/or monitor access as required and to guarantee compliance with access restrictions.

AIP (Archival Information Package): Within the OAIS Reference Model, the AIP is the managed, archival package of digital objects (including both content and metadata) that is stored and preserved by the OAIS system. The AIP is composed of the SIP plus any additional metadata related to any archival processes that were undertaken. See also: SIP, DIP. [1]

API (Application Programming Interface): A collection of computer subroutines published in such a manner that other software can easily invoke the subroutines. API is often synonymous with software library, subroutine library, or module. An API will contain subroutines, functions, methods, and/or classes.[2]

Appraisal: Process of deciding whether or not materials have enduring research value and should therefore be retained. The term selection is also sometimes used for this process.[3]

Archivematica: A comprehensive open source digital preservation software system that complies with the ISO-OAIS functional model.[4]

Archivists' Toolkit (AT): An "open source archival data management system to provide broad, integrated support for the management of archives." It is the result of a collaboration of the University of California San Diego Libraries, the New York University Libraries and the Five Colleges, Inc. Libraries. [5]

Arrangement: The process of organizing materials with respect to their provenance and original order, to protect their context and to achieve physical and intellectual control over the materials.[6]

---

[1] Source: CCSDS Recommendation for an OAIS Reference Model, pg 1-7.

[2] Source: http://en.wikipedia.org/wiki/Application_programming_interface)

[3] Source: http://www.archivists.org/glossary/term_details.asp?DefinitionKey=3

[4] Source: http://archivematica.org/wiki/index.php?title=Main_Page

[5] Source: http://www.archiviststoolkit.org

[6] Source: http://www.archivists.org/glossary/term_details.asp?DefinitionKey=294

Artifactual file: The original file, a copy of which is then processed in working/preservation storage - likely to have its own preservation policy

Audit trail: A means of tracking all the interactions with records within an electronic system so that any access to the system can be documented as it occurs for the purpose of preventing unauthorized actions in relation to the records as well as determining if relevant policies and procedures were followed or, if not, why they were not followed.

Authenticity: The quality of being genuine and free from tampering as well as being what it professes in origin or authorship.[7]

Axiell CALM (Computerisation for Archives, Libraries and Museums): Archives collection management software developed by Axiell. Widely used in the UK.[8]

Catalog: see Finding Aid.

Checksum: A unique numerical signature with a fixed, small length, derived from a file. Used to verify that two copies of a file are identical. Also referred to as a hash value.[9]

Content: The intellectual substance of a document, including text, data, symbols, numerals, images, and sound.[10]

Context: The organizational, functional, and operational circumstances surrounding materials' creation, receipt, storage, or use, and its relationship to other materials. [11]

DACS: Describing Archives: A Content Standard. An output-neutral set of rules for describing archives, personal papers, and manuscript collections, and can be applied to all material types. It is the U.S. implementation of international standards (i.e., ISAD(G) and ISAAR(CPF)) for the description of archival materials and their creators.[12]

Description: The creation of an accurate representation of a unit of archival material by the process of capturing, collating, analyzing, and organizing information that serves to identify archival material and explain the context and records system(s) that produced it.[13]

---

[7] Source: http://www.archivists.org/glossary/term_details.asp?DefinitionKey=9

[8] Source: http://www.axiell.co.uk/calm

[9] Source: http://en.wikipedia.org/wiki/Checksum

[10] Source: http://www.archivists.org/glossary/term_details.asp?DefinitionKey=627

[11] Source: http://www.archivists.org/glossary/term_details.asp?DefinitionKey=103

[12] Source: http://www.archivists.org/governance/standards/dacs.asp

[13] Source: InterPARES 2 Project, Terminology Cross-domain Task Force, pg 5.

Discovery model: Combination of environment, tools, and services which enables users to identify and locate resources of interest to them. At its most basic level, it requires publication or other dissemination of information, and increasingly it also implies on-line availability together with interactive search facilities.

Disk image: A single file or storage device containing the complete contents and structure representing a data storage medium or device, such as a hard drive, tape drive, floppy disk, CD/DVD/BD, or USB flash drive, although an image of an optical disc may be referred to as an optical disk image.[14]

Dissemination request: A request made of a repository or archive by a user for digital objects or metadata about them.[15]

DIP (Dissemination Information Package): The package of digital object(s) and metadata that is produced or retrieved by an OAIS system as a result of a dissemination request. See also: SIP, AIP. [16]

Donor: This term is used to denote any person or organization transferring material to an archival institution. The material may be a donation, purchase or deposit (indefinite loan). The term donor is used for convenience to imply any of these scenarios.

DRAMBORA (Digital Repository Audit Methodology Based on Risk Assessment): A methodology for self-assessment, encouraging organizations to establish a comprehensive self-awareness of their objectives, activities and assets before identifying, assessing and managing the risks implicit within their organization. Developed jointly by the Digital Curation Centre (DCC) and Digital Preservation Europe (DPE).[17]

DROID (Digital Record Object Identification): A software tool developed and distributed by the National Archives of the UK to that uses the PRONOM registry to automatically identify file formats.[18]

EAD (Encoded Archival Description): A non-proprietary de facto standard for the encoding of finding aids for use in a networked (online) environment.[19]

Emulation: The reproduction of the behavior and results of obsolete software or systems through the development of new hardware and/or software to allow execution of the old software or systems on future computers.[20]

---

[14] Source: http://en.wikipedia.org/wiki/Disk_image

[15] Source: CCSDS Recommendation for an OAIS Reference Model, pg 4-11.

[16] Source: CCSDS Recommendation for an OAIS Reference Model, pg 10.

[17] Source: http://www.repositoryaudit.eu/about/

[18] Source: http://droid.sourceforge.net/

[19] Source: http://www.archivists.org/saagroups/ead/aboutEAD.html

[20] Source: InterPARES 2 Project Book: Glossary, pg 20.

Fedora (Flexible Extensible Digital Object Repository Architecture): Originally developed by researchers at Cornell University as an architecture for storing, managing, and accessing digital content in the form of digital objects inspired by the Kahn and Wilensky Framework. Fedora implements the Fedora abstractions in a robust open source software system.[21]

Finding aid: A description of records that gives the repository physical and intellectual control over the materials and that assists users to gain access to and understand the materials.[22]

FITS (File Identification Tool Set): Identifies, validates, and extracts technical metadata for various file formats and combines their results into a single XML file. It wraps several third-party open source tools, normalizes and consolidates their output, and reports any errors. FITS was created by the Harvard University Library Office for Information Systems for use in its Digital Repository Service (DRS).[23]

File viewer: Application software that presents the data stored in a computer file in a human-friendly form. The file contents are generally displayed on the screen, printed, or read aloud using speech synthesis.[24]

Forensic disk image / forensic copy: A complete sector-by-sector copy of the source medium and thereby perfectly replicating the structure and contents of a storage device.[25]

Hybrid collection: A collection consisting of both born-digital and paper-based materials.

Hydra: A multi-institutional collaboration to build a common, open source framework for multi-function, multi-purpose, repository-powered applications. As symbolized by its name, Hydra's architecture reflects a "one body, many heads" design: a robust digital repository (the body) anchors feature-rich applications (the heads), tailored to content-, domain- and institution specific needs and workflows. Hydra's technical framework features the Fedora Repository on the back end, with a front end comprising Ruby on Rails, Blacklight, Solr, and a suite of web services.[26]

Hypatia: A Hydra application (Fedora, Hydra, Solr, Blacklight) that supports the accessioning, arrangement / description, delivery and long term preservation of born digital collections. By using a common set of software tools and APIs, Hypatia will also have features related to access, delivery, authorization, and preservation.[27]

---

[21] Source: http://www.fedora-commons.org/about

[22] Source: http://www.archivists.org/glossary/term_details.asp?DefinitionKey=66

[23] Source: http://code.google.com/p/fits/

[24] Source: http://en.wikipedia.org/wiki/File_viewer

[25] Source: http://en.wikipedia.org/wiki/Disk_image

[26] Source: https://wiki.duraspace.org/display/hydra/The+Hydra+Project

[27] Source: https://wiki.duraspace.org/display/HYPAT/Home

Ingest: The act of moving a submission information package (SIP) into a digital repository as an archival information package (AIP). Ingest also refers to a specific OAIS entity that contains the services and functions to perform this activity.[28]

Institution: The term used within this paper to describe a collecting repository, record office, or other institution undertaking stewardship of archives.

ISAD(G) (General International Standard Archival Description): A standard to provide general guidance for the preparation of archival descriptions. It is to be used in conjunction with existing national standards or as the basis for the development of national standards. Created by the International Council on Archives (ICA) Sub-Committee on Descriptive Standards (CBPS) (second edition, published 2010).[29]

Logical copy / Logical image: A copy of specific files made from a storage device, retaining their hierarchical organization within directories or folders. The full path of each file is recorded. Deleted files and un-partitioned space are not copied.

METS (Metadata Encoding and Transmission Standard): A standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed as an XML schema.[30]

Migration: The process of converting records to newer formats in order to maintain their compatibility with a newer generation of hardware and/or software computer technology, while leaving intact their intellectual form.

Normalization: The process of creating and/or storing digital documents or other digital objects in a limited number of standardized data or file formats.[31]

OAIS (Open Archival Information System): 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. (ISO 14721:2003)[32]

Open source: A development strategy wherein the source materials of an end product are made available. The term is most commonly used in collaborative software development when source code of an application is made available with an application for others to change or improve upon.

Original order: The organization and sequence of records established by the creator of the records.[33]

---

[28] Source: CCSDS Recommendation for an OAIS Reference Model, pg 11.

[29] Source: http://www.ica.org/7102/public-resources/isadg-general-international-standard-archival-description-second-edition.html

[30] Source: http://www.loc.gov/standards/mets/

[31] Source: InterPARES 2 Project Book: Glossary, pg 20.

[32] Source: CCSDS Recommendation for an OAIS Reference Model, pg 1-11.

[33] Source: http://www.archivists.org/glossary/term_details.asp?DefinitionKey=69

PAIMAS (Producer-Archive Interface Methodology Abstract Standard): Standard that covers the first stages of the ingest process defined by OAIS. Identifies and provides a structure for the interactions which take place between an information producer and a deposit archive. (ISO 20652:2006)[34]

Physical control: The function of tracking the storage of records to ensure that they can be located.[35]

PREMIS (Preservation Metadata: Implementation Strategies): Core preservation metadata model for organizing and thinking about preservation metadata, defined in the PREMIS Data Dictionary. Documentation includes guidance for local implementations.[36]

PRONOM: a web-based technical registry to support digital preservation services; developed by The National Archives of the United Kingdom. A proposal current underway seeks to bring PRONOM together with the Global Digital Format Registry Project to create a Unified Digital Formats Registry (UDFR). The DROID tool was created to use the PRONOM registry for format identification.[37]

Provenance: The relationship between records and the organizations or individuals that created, accumulated, and/or maintained and used them in the conduct of personal or corporate activity. See also: respect des fonds.[38]

Processing environment: The workspace where accessioning and arrangement and description is undertaken.

Quarantine space: A location where items can be held and isolated in order to mitigate the effect of any contaminants and prevent them from spreading to other materials. In a workflow with born-digital materials, this may involve keeping files on storage media not connected to a server until malware or viruses can be detected and removed.

RAD (Rules for Archival Description): Published by the Canadian Committee on Archival Description. Revised version released in 2008.[39]

Repository: Term used in this paper to refer to the digital repository, not as an alternative term to institution.

---

[34] Source: http://www.dcc.ac.uk/resources/standards/diffuse/show?standard_id=154

[35] Source: http://www.archivists.org/glossary/term_details.asp?DefinitionKey=978

[36] Source: http://www.loc.gov/standards/premis/tutorials.html

[37] Source: http://www.nationalarchives.gov.uk/PRONOM/

[38] Source: http://archives.un.org/unarms/en/unrecordsmgmt/unrecordsresources/glossaryofrecordkp.html

[39] Source: http://www.cdncouncilarchives.ca/archdesrules.html

Respect des fonds: The principle that the records created, accumulated, assembled, and/or maintained and used by an organization or individual must be kept together in their original order if it exists or has been maintained and not be mixed or combined with the records of another individual or corporate body.[40]

Ruby on Rails: An general, object-oriented open-source programming language. Ruby is tightly integrated with a web application framework called Rails. Ruby on Rails is part of the stack of technologies used in the Hydra and Hypatia projects.[41]

Series: A group of records based on a file system or maintained as a unit because the records result from the same function or activity, have a particular form, or have some other relationship resulting from their creation, accumulation, or use.[42]

Significant properties: Significant properties, also referred to as "significant characteristics" or "essence", are essential attributes of a digital object which affect its appearance, behavior, quality and usability. They can be grouped into categories such as content, context (metadata), appearance (e.g. layout, color), behavior (e.g. interaction, functionality) and structure (e.g. pagination, sections). Significant properties must be preserved over time for the digital object to remain accessible and meaningful.[43]

SIP (Submission Information Package): The data and metadata received into an OAIS-system at Accessioning as part of the ingest process. See also: AIP, DIP. [44]

Stabilization: Establishing a safe and secure digital environment for the long-term preservation and storage of electronic records.

TAPER (Tufts Accessioning Program for Electronic Records): A tool developed by Tufts University to create submission agreements for electronic records. Flexible enough to apply to many types of born-digital materials.[45]

TDR (Trusted Digital Repository): A standard developed by RLG and OCLC to define the characteristics of a sustainable digital archive that could serve large-scale, heterogeneous collections held by such research repositories as national libraries, university libraries, special collections, archives, and museums.[46]

---

[40] Source: http://www.archivists.org/news/custardproject.asp?prnt=y

[41] Source: http://rubyonrails.org

[42] Source Roe pg. 61

[43] Source: http://www.jisc.ac.uk/whatwedo/programmes/preservation/2008sigprops.aspx

[44] Source: CCSDS Recommendation for an OAIS Reference Model, pg 1-13

[45] Source: http://sites.tufts.edu/dca/about-us/research-initiatives/taper-tufts-accessioning-program-for-electronic-records

[46] Source: http://www.oclc.org/research/activities/past/rlg/trustedrep/

TRAC (Trusted Repository Audit and Certification): A set of criteria to facilitate the certification of digital repositories capable of reliably storing, migrating, and providing access to digital collections. Developed by RLG and the US National Archives and Records Administration of the United States.[47]

Virtual machine: A "completely isolated guest operating system installation within a normal host operating system". It is a software implementation of a machine (i.e. a computer) that executes programs like a physical machine.[48]

[47] Source: http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying-0

[48] Source: http://www.griffincaprio.com/blog/2006/08/virtual-machines-virtualization-vs-emulation.html