Persistent Digital Archives and Library System (PeDALS)

The Persistent Digital Archives and Library Sytem (PeDALS) was first implemented in 2008, and meant to continue for 18 months. Headed by the Arizona State Library, and counting with 4 other partners (Florida, New York, South Carolina, and Wisconsin), the project was so successful that it extended its life spam, and, in 2010 had the adherence of two other partners: Illinois and New Mexico.[1]

PeDALS consists of a curatorial and archive research project that makes efforts towards lower costs of digital preservation. PeDALS implements inexpensive "digital stacks," which are a storage network that can preserve a collection's integrity and authenticity. The project is funded by the Library of Congress's National Digital Information Infrastructure and Preservation Program (NDIIPP) and Institute for Museum and Library Services, and received, for its initial 18-month life spam, $2.25 million.[2] According to the Library of Congress, "the goal of PeDALS is to develop a shared curatorial framework for the preservation of digital public records such as agency publications and court records," all born digital.[3]

The Quarterly Newsletter of the State Library and Archives of Florida summarized in 2008 the goals and structure of PeDALS:

> When the system is operational, it will allow state government agencies to automatically retrieve and transfer massive quantities of electronic records into the storage system. Digital copies of these records will be stored across the nation at different points in the network, which will protect them in case of a catastrophic disaster in one state. The preservation of these records, ordinarily something too large and time consuming to do manually, will be done automatically.[4]

PeDALS follows a curatorial rationale that complies with the OAIS (Open Archival Information System) standards. Its requirements for a Submission Information Package (SIP) are relatively simple so that the offices of origin of materials can easily meet them, and include in the SIP roughly the record itself, added with key metadata. At the point of ingest, or dropbox, first place in which content is deposited, the record is scanned for viruses and errors, and metadata is checked to meet the specifications established by the repository. According to PeDALS:

---

[1] Library of Congress. Digital Preservation. News & Events. Apr 1 2010.
<http://www.digitalpreservation.gov/news/2010/20100401news_article_PEDALS_partne rs.html>

[2] State Library and Archives of Florida, January-March 2008 Newsletter.
<http://dlis.dos.state.fl.us/newsletter/article.aspx?articleID=1136&newsID=1023>

[3] Library of Congress. Digital Preservation. News & Events. Apr 1 2010.
<http://www.digitalpreservation.gov/news/2010/20100401news_article_PEDALS_partne rs.html>

[4] State Library and Archives of Florida, January-March 2008 Newsletter.
<http://dlis.dos.state.fl.us/newsletter/article.aspx?articleID=1136&newsID=1023>

A significant part of the ingest process involves mapping metadata received with the materials to the core metadata. Developing the core metadata dictionary includes:

> • Identifying and defining the elements that support administration, discovery, and preservation functions;
> • Establishing the requirement level for the elements;
> • Testing and refining the set using metadata received with records and publications;
> • Testing and refining the set to ensure they support the system's administrative, discovery, and preservation functions.[5]

Its metadata system includes content-based metadata (which can either be submitted with the original material, or inferred from it) and technical, administrative and preservation metadata (originated during its process).

Because of the use of an automated and integrated workflow to process collections of digital records and publications, curators work with rules, and not with the records themselves. Before transfer rules are launched, however, the curator must validate the SIP to ensure that it conforms to the repository's specifications. PeDALS also uses the New Zealand Metadata Extractor (MetaExtractor) to create software format, version, file type, and mime type. MetaExtractor programmatically extracts preservation metadata from different file formats.

PeDALS uses an automated middleware system of rules and logics that finally processes its records, after the ingest process. The middleware, BizTalk, is – at its core – an XML message queue that receives objects, converts and performs logic on them, and sends them out as an Archival Information Package (AIP).

The AIP uses an XML schema, which includes the submitted record, original metadata from the office of origin (OOO), additional metadata generated by the repository, and specific digital signatures. PeDALS provides accession to their register database, in addition to rules to ingest AIP into a LoCKSS (Lost of Copies Keeps Stuff Safe) cluster. This cluster is believed to ensure a longer preservation of the record, as copies of the AIP are geographically dispersed, and stored in different, independent servers. "Each participating state [has] a PeDALS system of its own to ensure that each state's records will be kept completely separate and data will not be intermixed."[6] This multi-copy structure allows different servers to be administered differently, further ensuring the records' safety. LoCKSS uses a redundant array of inexpensive servers, and provides automatic integrity checks and error corrections, besides preserving the record's bitstream.

Although the AIP is the official copy of the record (used for storage and management), it is not user-friendly, and the repository generates a Dissemination Information Package (DIP) in an accessible format that also protects the record's integrity. The DIP follows specific publishing rules established by the repository, and does not include administrative or preservation metadata. When not confidential, the

---

[5] PeDALS: Core Metadata. <http://www.pedalspreservation.org/Metadata/Default.aspx>
[6] Library of Congress. Digital Preservation. News & Events. Feb 1 2008.
<http://www.digitalpreservation.gov/news/2008/20080201news_article_az.html>

DIPs are put on public webserver in a format easily supported by most commonly used browsers. According to PeDALS:

> The DIPs used for public access will be stored on a separate system. Similarly, the accessions register database will be protected by firewalls and have limited access, while the metadata for publicly accessible records will be stored in a separate database to avoid the possibility of inferring the content of confidential or restricted records from the metadata.[7]

PeDALS holds all its records' information at its Administrative Catalog (or accessions register database), which lives in a different, independent server. This database informs records' metadata, and is not publicly available. The schema below illustrates the architecture of the Arizona PeDALS:

_____
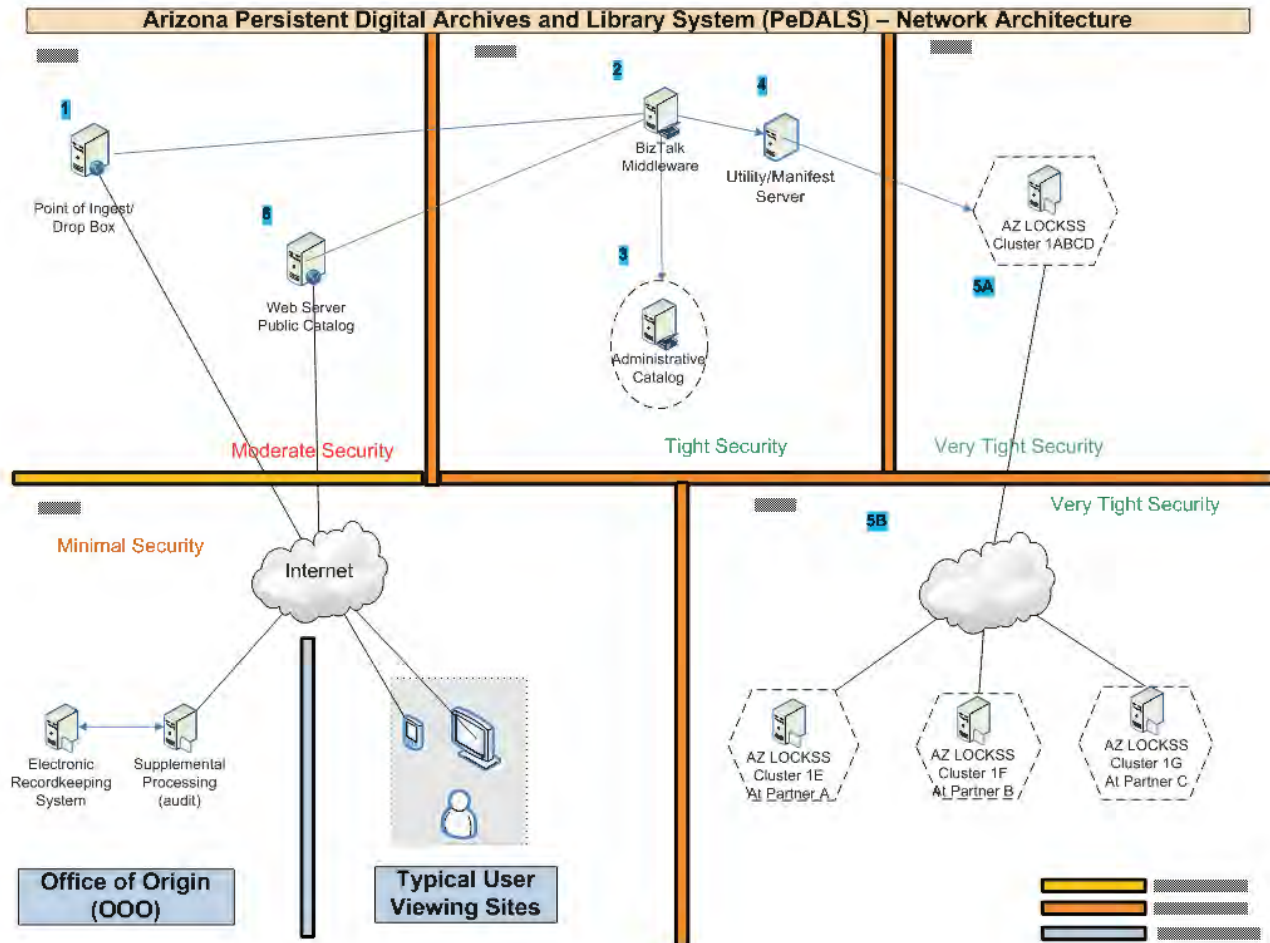
[7] PeDALS: Digital Stacks. <http://pedalspreservation.org/About/stacks.aspx>

Since heavily relying on partnership (counting, today, with a total of seven state partners), PeDALS aims to build a community of shared practice, in order to promote the interoperability of digital repositories for ongoing development of the system (through rules of best practices), and to avoid redundant work. This way, the repositories ensure the safety and integrity of digital records (through LoCKSS technology) and push for the development and dissemination of inexpensive preservation strategies, while still maintaining their administrative independence.

Whereas PeDALS pays special attention to the usability of its dissemination packages' records, it does not offer a user interface. Instead, it distributes the packages back to the office of origin, and/or to a specific site on the webserver. PeDALS appears to tackle on the issue of digital preservation of public electronic records that have better usage in its original format (i.e., digital) than in its analog equivalent.

Additional Webography

"Curating the Digital Past: Lessons from the PeDALS Project" in University of Arizona's College of Social and Behavioral Sciences: http://web.sbs.arizona.edu/college/node/506

Library of Congress, Digital Preservation Partners:
http://www.digitalpreservation.gov/partners

Library of Congress, Digital Preservation Resources:
http://www.pedalspreservation.org/Resources/PeDALS_Pilot_Architecture%20(v2%20si mple).pdf

PeDALS: http://www.pedalspreservation.org

PeDALS's Email Extractor: http://pedalsemailextr.sourceforge.net/