

IBM Haifa Research Lab's Preservation DataStores technology helps preserve digital data and interpretations of the objects represented by those bits for decades

## Saving your Digital Data for Ages

### What is Preservation DataStores?

We can read and interpret the Dead Sea scrolls written almost 2000 years ago, but we cannot do the same with data generated 20 years ago on a 5.25 inch floppy disk. Ironically, as the world becomes digital, we may be entering a digital "Dark Ages" in which business, public and personal assets are in ever greater danger of being lost. We are capable of storing digital bits, and can interpret them, as long as the environment doesn't change. But what current technologies can ensure storage and interpretation over long periods of time?

Long Term Digital Preservation (LTDP) environments facilitate sustaining the understandability and usability of digital information, even when technologies for computer hardware, operating systems, data management products, and applications are replaced with newer ones.

**Preservation DataStores (PDS)** is an innovative technology that supports digital preservation environments, ensuring data usability and integrity over long periods of time. Past experience shows that digital preservation is not incidental. Unless you care for the data in a proactive manner, its understandability and usability will be lost in about 10 years. PDS automates the processes and management needed to care for the data and enables its preservation.

PDS serves as an infrastructure component of CASPAR, a European Union project that is building a framework to support the end-to-

end preservation lifecycle for scientific, artistic, and cultural information.

### Support Logical Preservation

#### The Challenge

Traditional archival storage primarily considers bit preservation, if it considers preservation issues at all. But, preserving the bits is not enough! You also, need to ensure logical preservation to help interpret the data down the line.

#### The Solution

The Open Archival Information System (OAIS), an ISO standard since 2003 ([ISO 14721:2003 OAIS](https://www.iso.org/standard/54467.html)), is a core standard for digital preservation systems. The basic artifact in OAIS is called an archival information package (AIP).

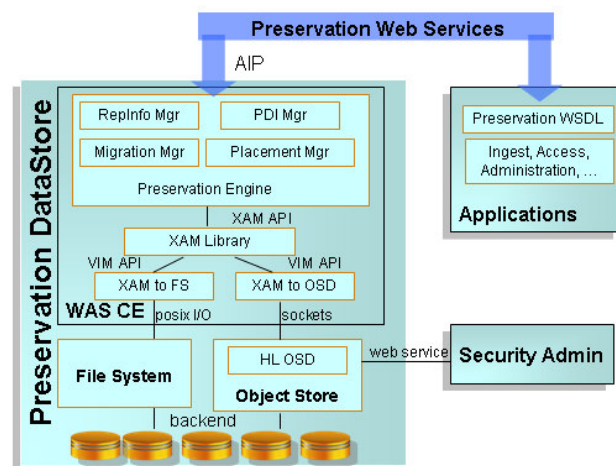


Figure 1: Preservation DataStores architecture

PDS is a novel storage architecture that has built-in support for logical preservation based on OAIS. PDS generates AIPs for the raw data and encapsulates the various AIP parts, so they are inseparable during the migration processes and when accessing the data in the future. PDS decreases the data transfer between the applications and the storage by offloading data intensive functions such as fixity (i.e., integrity) computations and transformations to the storage—thereby simplifying storage-related management.

PDS is composed of a layered architecture based on open standards, along with the OAIS, XAM and OSD standards. It complies with the general design principle of preservation systems that employ open standards wherever possible.

---

## PDS Integrated with ECM

### The Challenge

Customer requirements and recent compliance legislation, such as HIPAA and the Sarbanes-Oxley Act, which demand long-term data viability, have brought about the need for Enterprise Content Management (ECM) systems to support long term digital preservation.

### The Solution

PDS Preservation Engine can be easily integrated with an ECM system providing an ECM with OAIS-based logical preservation support. The Preservation Engine maps the AIP to the ECM object model utilizing its advantages as well as ensuring AIP encapsulation and co-location. It then runs on top of storage that supports bit preservation.

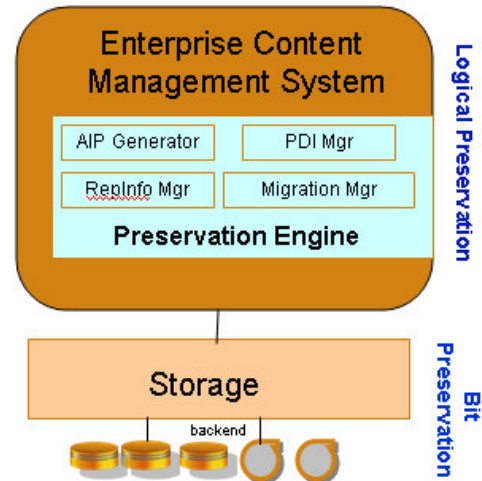


Figure 2: PDS integrated with ECM architecture

---

## LTDP Assessment Tool

The LTDP **assessment tool** can determine the readiness and maturity level of an organization's repository intended to preserve data for the long term. For example, if a company has a repository of digital goods, the assessment tool will indicate whether these items are being preserved correctly, whether the processes and technologies are appropriate, and so forth. The assessment tool is based on compliance with an audit checklist that was developed by the digital preservation community and is being transformed into a certification standard. An intuitive web tool takes the user through a set of questions and provides a score to assess the quality of the long term repository.

### Contact us for more information

Simona Cohen, [simona@il.ibm.com](mailto:simona@il.ibm.com)  
Dalit Naor, [dalit@il.ibm.com](mailto:dalit@il.ibm.com)  
Michael Factor, [factor@il.ibm.com](mailto:factor@il.ibm.com)  
<http://www.haifa.ibm.com/projects/storage/dastores/index.html>