

Building a Framework for Applying OAIS to Distributed Digital Preservation

Katherine Skinner (Educopia Institute), Eld Zierau (The Royal Library of Denmark), Matt Schultz (MetaArchive Cooperative)

Worldwide, digital preservation initiatives are engaging highly distributed methodologies, infrastructures, and organizational apparatuses in order to achieve the reliable persistence of digital content. Examples include MetaArchive Cooperative, LuKII, Chronopolis, DuraCloud, UC3 Merritt, the Danish Bit Repository, Internet Archive, COPPUL, iRODS, LOCKSS, Archivematica, DPN, and many others.

This first generation of preservation systems and approaches has emphasized the importance of such factors as content replication, geographic distribution, and coordination, (e.g. through infrastructure heterogeneity and organizational diversity for ensuring the longevity of digital objects.)

Although the use of distribution is common within the preservation field, there is not yet an accepted definition for “distributed digital preservation” (DDP). As the preservation field has matured, the term “distributed digital preservation” has been applied to myriad preservation approaches, from initiatives that intentionally build distributed infrastructures as essential components of their preservation models (e.g., the examples named above) to other initiatives that merely engage in back-up measures for their preservation repository by storing digital copies of their managed content off site.

Existing DDP approaches lack common vocabulary and conceptual frameworks for building effective, reliable, and auditable distributed preservation environments. Such agreed-upon terminology and theoretical models would help initiatives to describe and compare their infrastructures and operations. They would also increase understanding and awareness of the distributed preservation process both by practitioners and by organizations seeking solutions. Finally, they would provide auditors (including self-auditors) with a crucial foundation for assessing the reliability of various distributed approaches.

The Educopia Institute, the Royal Library of Denmark, the Library of Congress, Chronopolis, DuraSpace Foundation, MetaArchive Cooperative, LOCKSS, Data-PASS, California Digital Library, Archivematica, and Internet Archive are developing a Framework to identify, define, and provide a documented model for the range of relationships and interactions that occur in Distributed Digital Preservation (DDP) environments. This group defines DDP as the use of replication, independence, and coordination to address the known threats to digital content to ensure its accessibility through time.

The intent of this international alignment effort is to effectively document the practice of distributed digital preservation through 1) articulating a common set of concepts and terminology to help build a common understanding of DDP practices, 2) describing scenarios for placing OAIS functional entity components *across* distributed environments (including those that entail the coordination/collaboration of multiple organizations), and 3) describing the roles/responsibilities that might be involved in replicating/preserving digital content in technically and/or organizationally distributed environments.

Thus far, this group has authored a white paper describing the range of trends and practices in DDP and how they compel an elaboration upon OAIS. The paper also identifies challenges that this endeavor might face as it moves forward, and puts forward a series of recommendations for progressing toward a formal framework for a DDP environment. Development of this framework will be based on comprehensive interviews with and feedback from distributed digital preservation groups worldwide.

In order to advance the initiative, the project group needs to expand to include other international exemplars of DDP. Aligning our perspectives and successfully articulating and documenting DDP concepts will help to ensure that DDP efforts can measure and demonstrate their reliability, both technically and organizationally. It will also help archival institutions to understand the role and value of heterogeneity, geographic distribution, and organizational diversity in the long-term care of digital objects.