**The Digital Preservation Network: Architecture, Processes and Policies**

David Minor
Chronopolis Program Manager
UC San Diego Library

Tom Cramer
Chief Technology Strategist
Stanford University

James Simon
IT Architect
Stanford University

Scott Turnbull
APTrust Technical Lead

The Digital Preservation Network (DPN) was formed in 2012 to ensure that the scholarly record and output of scholarly research are preserved for future generations. DPN uses a federated approach to preservation, bringing together extant preservation systems to leverage existing expertise and infrastructure. The higher education community has already created digital repositories to provide long-term preservation and access. Replicating multiple dark copies of these collections in diverse nodes, DPN will protect against the risk of catastrophic loss due to technology, organizational or natural disasters.

Depositors will add digital content to DPN via a DPN Node. This content is then replicated to other DPN Nodes, which together form a network of heterogeneous, secure, trustworthy digital archives. Each DPN Node is operated under diverse geographical, organizational, financial, and technical regimes. Bit auditing and repair functions ensure the fixity of content over time. Intellectual property agreements ensure the succession of rights to use of the content through the Network in the event of dissolution or divestment of content by the original depositor and/or archive.

This presentation will present technical details of DPN, including:

**Architectural Diagrams of relevant infrastructure at each of the initial five DPN Nodes:**

* HathiTrust
* Stanford Digital Repository
* Chronopolis
* Texas Digital Library
* Academic Preservation Trust

**System Components of the DPN Nodes**:

* Messaging: how nodes communicate in the Network.
* Registry: method used for tracking items in the Network.
* Auditing: how fixity is maintained in the Network.
* Encryption: Decisions made on how content is to be protected during transmission and storage.
* Reporting: How successful and unsuccessful completions of events and errors are transmitted in the network.

 **Data Elements in DPN**:

* + Content Packaging: use of BagIt and Bagit profiles.
	+ Identifiers: how objects are referenced within the Network.

We will also present a typical flow of Primary Interactions in DPN, explaining the important details. For example:

* Negotiation with a DPN Node (a “First Node”) to ingest content. Initial preservation actions taken, agreements signed.
* Replication of content from First Node to other DPN Nodes (“Replicating Nodes”), who function as dark preservation archives for the content.
* Bit auditing and repair functions that continuously monitor the integrity of copies across all nodes, and will detect and repair corruption, errors, and data loss on an ongoing basis.
* Restoration of content from the Replicating Nodes in the event the First Node’s copies are lost or corrupted.
* Redistribution of preserved content as Nodes enter and leave the Network, ensuring continuity of preservation services over time
* Preservation and management of sufficient identity, administrative (including legal and contractual), technical, and descriptive metadata to enable the future brightening of content after succession, independent of any contributions or participation by the First Node or original Depositor.

The presentation will also touch on the policy and operational issues that have emerged as DPN has taken shape, and that are particular to a distributed network of heterogeneous preservation repositories. These include:

* what is a meaningful standard for trustworthy operation of the network, and for each node in the network?
* what is the best way to enable content to be brightened, after the presumed dissolution of the depositor and original archive?
* where are diverse technologies and preservation approaches in the architecture best employed, and where does a consolidated or shared approach make sense?
* what contractual framework makes the most sense for business operations as well as succession of intellectual property rights?
* what is the best way to manage encryption such that the preservation, security and contractual concerns are all addressed?
* how best to manage content versioning and content replacement across the network?