

Exploring Preservation Services in a Mediated Peer to Peer Federation



PASIG - Oxford
September 2017

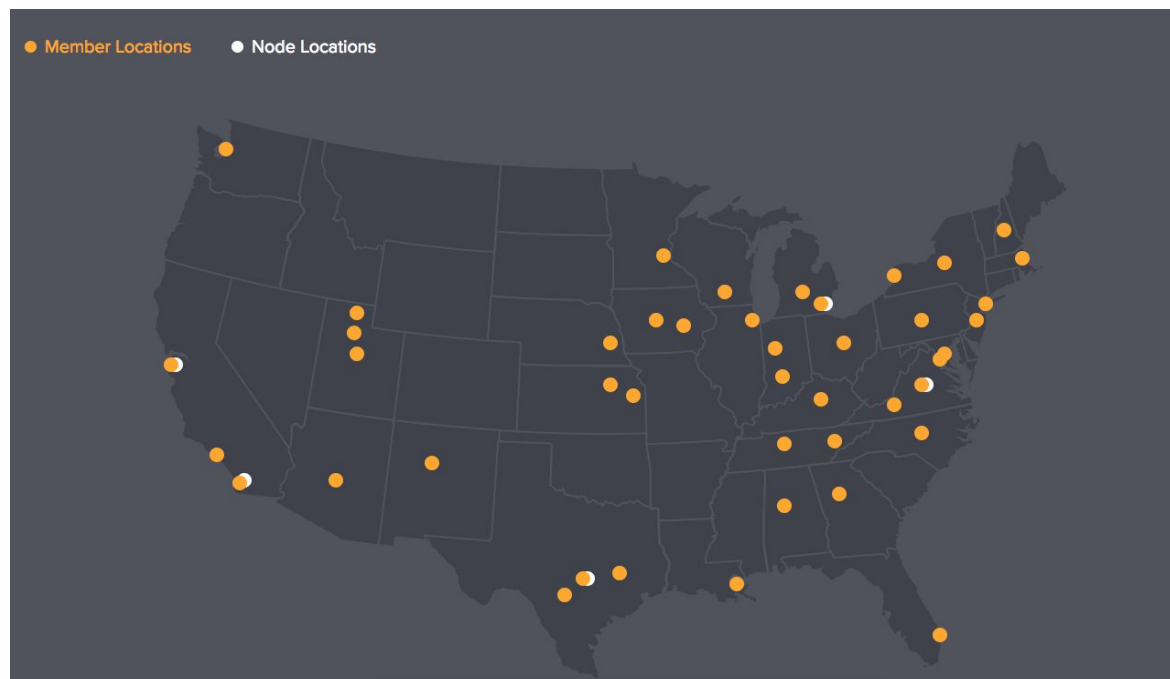
David Pcolar - Technical Officer

dave@dpn.org

<http://www.dpn.org>

What is DPN?

- A membership organization dedicated to the long-term preservation of the scholarly record
- A federation of cooperating repositories (nodes) supporting dark preservation
- A set of services and agreements that, at the time of deposit, commit to ensure that scholarly assets remain available for restoration by the academy



<http://dpn.org/about>

Sustainability and Community Involvement

“The fundamental problem is not storing bits safely for the long term, it is paying to store bits safely for the long term.”
- David Rosenthal

“Traditional business models relied on planning, which in turn assumes either that the future will be like today or that we are good at forecasting. Neither has proved true.”
- Peter Cappelli



Core Assumptions

- DPN is a federated preservation network of *independent archives*
- DPN has, as much as practicable, *independent implementations* at each service node
- DPN service nodes follow IT security best practices
- DPN provides for replication of content between nodes
- Deposited objects are preserved in dark preservation storage
- Nodes support preservation functions
 - Data provenance ¹
 - Periodic fixity
 - PREMIS event logging ²

¹ https://en.wikipedia.org/wiki/Data_lineage#Data_Provenance

² <http://www.loc.gov/standards/premis/v3/>

Who is DPN?

- An non-profit LLC (Limited Liability Corporation), operating under Internet2
- Mary Molinaro - Executive Director
- Dave Pcolar - Technical Officer
- Steve Morales - Business Development consultant
- DPN Board
 - University Presidents and Chancellors, CIOs, Librarians and industry leaders.
 - Michael McRobbie, Indiana University President, chair since inception
- Member Committees and Groups



DPN
Staff

Program
Directions

Metadata
Standards


Membership
and Finance

Heavy
Users

Policy

Technical
Advisory

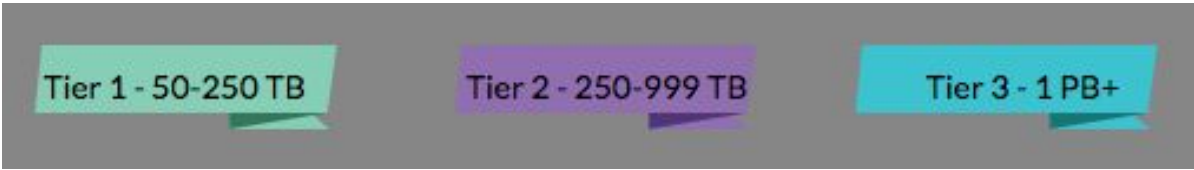
<http://dpn.org/>



Heavy
Users

Community Driven Heavy Users Committee Recommendations

1. Support a **new type of node scenario of Peers offering long-term storage**
2. Develop a **pilot project to explore scenarios** for the heavy user community.
3. Adopt 3 tiers of heavy users



Tier 1 - 50-250 TB

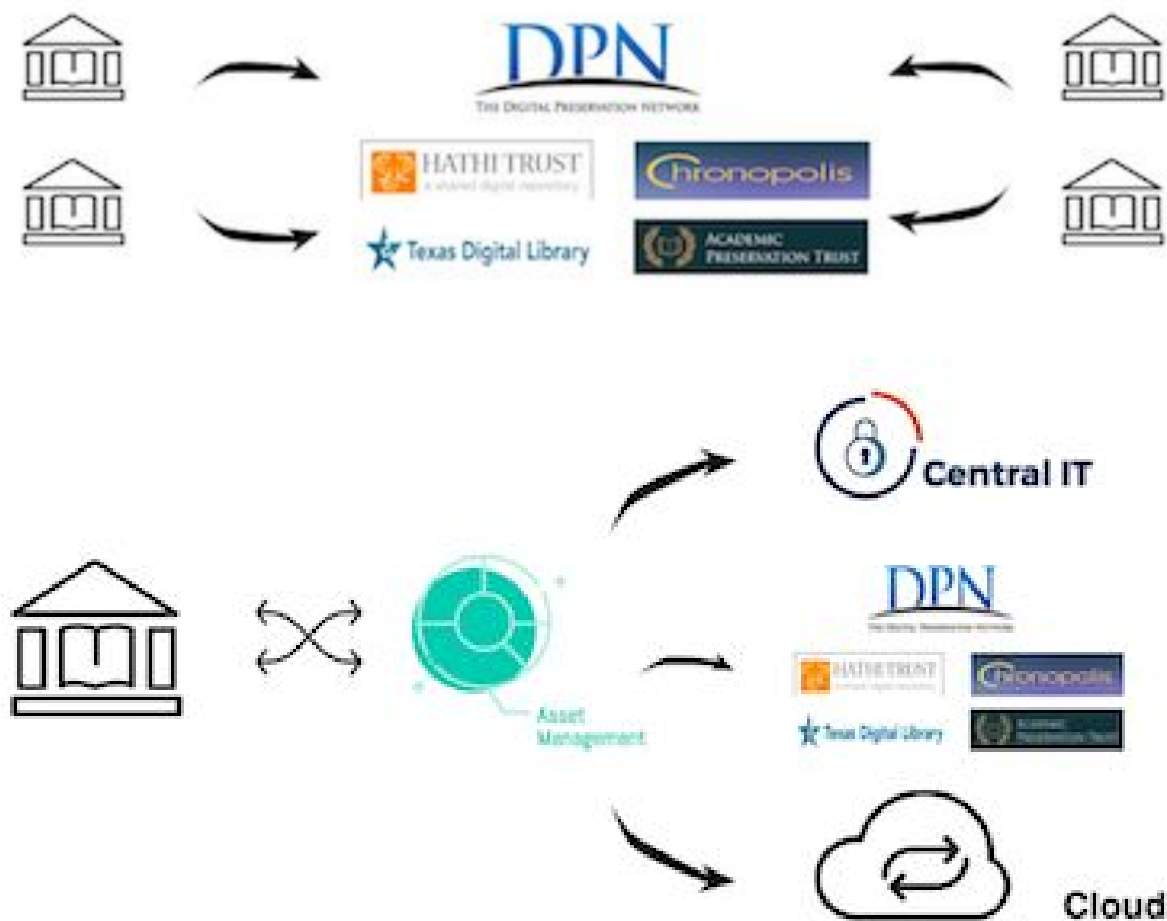
Tier 2 - 250-999 TB

Tier 3 - 1 PB+

4. **Explore service options** that support fewer copies.
5. Charge a Working Group to **explore large-scale data movement tools**.
6. Charge a Working Group to **develop financial and governance structures**.

<https://iu.box.com/v/DPN-HUC-Recommendations>

Preservation Modeling



P2P as a Preservation Architecture

“[#DigitalPreservation](#)’s future isn't a central repo, but leveraging distributed assets.” ³

DPN’s core Preservation Model continues to be the main service offering.

However...

Members desire to incorporate DPN into a broader institution centered model:

- where DPN’s 3-copy strategy might be overkill
- where institutional requirements drive preservation policy
- where members have in-kind resources that they can commit

DPN P2P Pilot

Initial Steps

- 1. Establish a high level roadmap of the pilot process and goals**
- 2. Define distinct areas of exploration**
 - a. Policy
 - b. Membership/Finance
 - c. Technical
- 3. Identify and report known constraints**
- 4. Pair partners for testing**



Peer Scenarios

Peer Mediated Ingest

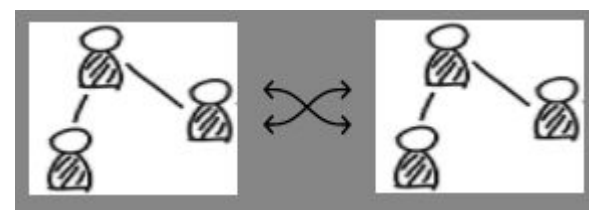
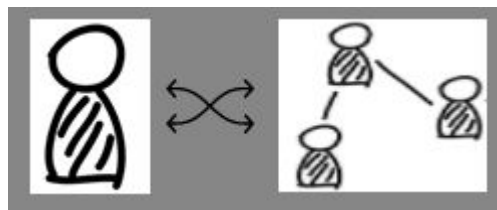
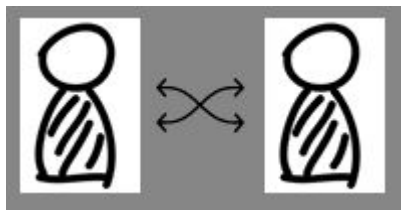


Alternate Ingest paths into DPN

- A workflow is created to ingest content via a peer landing zone
- Integrate use of large scale transfer tools
- Transfer content out of band of DPN ingest registration
- Simplify content registration in DPN

Peer Scenarios

Peer to Peer



Exploring new transaction models

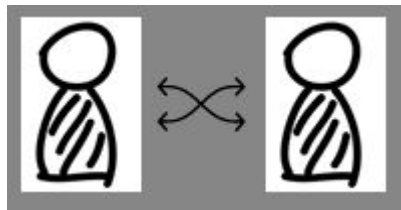
- Negotiate common transfer and deposit agreements
- Identify DPN's role in the transactions
- Establish workflows for content registration and metadata exchange
- Evaluate sustainability of the models



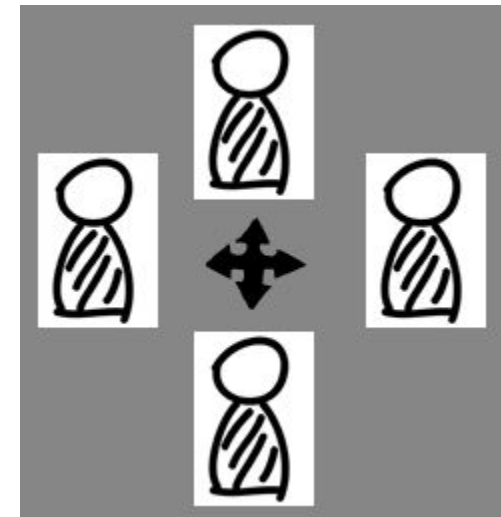
DPN's Value Add

- **Depositor and Contributor Agreements**
 - Negotiating common terms
 - Establishing DPN as a 'clearing house'
 - Consistent Level of Service and exit strategies
- **Independent Content Tracking and Assurance**
 - Transfer and storage verification
 - Transaction auditability
- **Workflow and Metadata Guidance**
 - Preservation metadata 'best practices'
 - Digital preservation workflow curriculum

Preservation Architecture Value Chain



- Emphasizes the importance of collaboration
- Effective partners align and complement skill sets
- Equitable resource commitment and consumption
- Services may not be linear, encouraging many:many relationships
- Trust and commitment are inherent in the relationships



Rethinking Digital Preservation

- Preservation is a game of probabilities
- Selective actions mitigate risk of loss
- Metadata creation, policies, and procedures are foundational elements
- Diverse technical architectures reduce risk

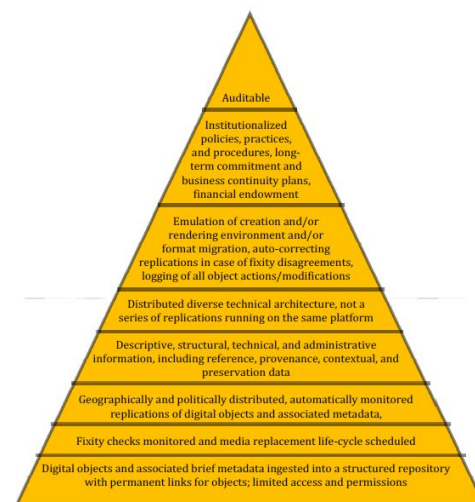


Figure 1: Hierarchy of Digital Preservation Needs [1, 2]

“Sustainable preservation strategies are not built all at once, nor are they static.”

["Sustainable Economics for a Digital Planet"](#)

[Rethinking Digital Preservation, Wilson](#)

Joint Preservation Code of Ethics

The mission of our shared work is rooted in a set of core values. These core values inform and direct our collective work:

Collaboration - working together and with others to preserve digital assets

Affordability - pricing to sustain and grow services

Availability - access to a variety of services for all, regardless of size or technical ability

Diversity - in our partners, collections, and technical solutions

Stewardship Continuity - by working together to help identify new locations for content that is at risk when one of our organization's stewardship cannot continue.

http://bit.ly/Preservation_codeof_Ethics



Thank You!

dave@dpn.org

P2P References:

- [Long-term Digital Archiving Based on Selection of Repositories Over P2P Networks](#)
- [Peer-to-peer data preservation through storage auctions](#)
- [Network Neutrality \[Paid peering: pricing and adoption incentives\]](#)
- [Internet service provision and content services: paid peering and competition between internet providers](#)
- [KopperCoin – A Distributed File Storage with Financial Incentives](#)
- [P2P commercial digital content exchange](#)
- [Peer to peer data trading to preserve information](#)
- [Creating trading networks of digital archives](#)