

DPN Restoration Pilot

Overview

DPN is facilitating a Pilot during which Members with large data transfer and preservation needs can explore the processes and policies needed to support their activities. By capturing as much information as we can about preparing and ingesting materials and metadata, simulating the legal negotiation required, talking through how transactions would be facilitated in the DPN architecture, and thoughtfully considering other issues encountered along the way, DPN hopes to refine both its understanding of Member requirements as well as services to meet them.

The DPN Heavy Users Committee (HUC), along with DPN staff formulated and submitted a set of recommendations that identified important components of cultivating use of DPN by members with large-scale data preservation needs. Select committee members, along with DPN staff and members of the Technical Advisory Committee, evaluated the recommendations and selected a set of steps to test the feasibility of the recommendations.

This document is intended to guide this Pilot by providing a framework for communication among all participants and by suggesting specific discussion topics and measurement metrics for a final report.

Communication and Coordination

The Peer to Peer Pilot will be time and effort limited, with a focus on transaction building, execution, and validation. For the duration of the pilot, participants and select DPN committees and groups will agree to:

1. Conduct a bi-weekly call to discuss status and major issues and to compare notes. This discussion will help to expose common issues and help all of the participants to characterize the issues similarly.
1. Create a report at the end of the process that encapsulates the discussions and discoveries brought to light. The report will be a formative document for production DPN services and will also be used to generate specifications for future pilots.
2. Confer with technical contacts at Peer sites, Node representatives and DPN leadership to help insure smooth implementation of the pilot.

Framing Questions

The following questions are meant to guide the discussion between participating Peer Members who are working together, as well as the larger discussion. This should not be considered an exhaustive list of issues to explore and information to collect, but rather to illustrate the type of information that will be valuable to capture.

Questions for Participants to Consider

1. Related to transfer technology and data selection
 - a. Should peers use a common bagging and validation tool?
 - b. What are the size limits for bagged content?
 - c. Which large scale transfer protocols/mechanisms should be considered?
 - d. For the purposes of the pilot, will the transfers involve tool API access?
2. Related to content preparation and transfer
 - a. Why did you want to deposit this content to a peer?
 - b. How would you go about determining what content should go a peer and what may be deposited into DPN?
 - c. How long / how much effort did or would it take you to do the following:
 - i. select content?
 - ii. negotiate legal agreements for content transfer and retention?
 - iii. package content?
 - iv. transfer content?
 - d. What thinking and/or strategy went into devising your content packaging?

- e. What kind of feedback and/or confirmation of receipt did you want, need, and/or receive for transfer and storage?
- f. What specific parts of the process have strong potential to cause delay?
- g. What would you advise others who are preparing to transfer content?
- h. What documentation about content preparation, above and beyond that provided, was useful or would have been useful to you?

3. Related to DPN services

- a. Was your content packaging designed with retrieval/restoration in mind? How so?
- b. What did you encapsulate or consider encapsulating with the object to enable interpretation and understand the content?
- c. Did you look at the legal issues around depositing restricted intellectual property, and subsequent use?
- d. What special concerns arose from your individual institutional perspective?
- e. What kind of feedback and/or confirmation of receipt did you want, need, and/or receive about transfer into the *DPN* network?
- f. What sort of regular accounting reports would be useful?
- g. What kind of ongoing preservation integrity audit reports would be useful?

4. What did you learn from the process overall?

5. What changes would you recommend for future Pilots?

Questions related to Peer-DPN interactions

1. How much time and effort was required for
 - a. conducting initial discussions
 - b. Executing legal documents
 - c. Registering content into DPN
2. What issues or anomalies did you encounter in transforming supplied content to your specifications or in packaging it for DPN?
3. Did you consider contractual / agreements concerns particular to DPN as part of the transfer? What issues, concerns, or outcomes arose?
4. Did you consider future recovery of the content via DPN services?
5. What do you need from DPN to augment your reporting of peer/DPN held content?

Committee Assignments

https://docs.google.com/document/d/1QCzBtuOY_KRySznrvxcijmXzv0jAPWmFcBw5t8r2X4E/

Successful deployment of peer services will require policy, pricing, and technical changes to the DPN operational model. We envision some decision to be made prior to a pilot completion, while others will be ongoing and adaptive to Member experiences and needs. Where possible, initial decisions will be made by direct action and reviewed for effectiveness and compliance with DPN policies and governance.

• Membership & Pricing Committee

- The Membership & Pricing Committee is tasked with proposing membership models and evaluating pricing models.
- Membership Tasks
 - Determine membership models which decouple storage from membership
 - Tier membership levels
 - Validate need for different lengths of service
- Pilot tasks
 - What is the membership fee for Peer nodes?
 - Does membership cover a predetermined number of transactions?
 - Are there any per transfer/registration fees?

- **Technical Advisory Committee**

- This group provides input to the strategic agenda for preserving content and exploring new and different ways of delivering digital preservation.
- Provide input and prioritize development and integration efforts.
- Evaluate tools in use at member institutions and assess possibilities for integration with DPN workflows and processes.
- Recommend new tools and processes to implement sustainable practices.
- Pilot Tasks
 - Specification for Peer node tools
 - Specification for Node Proxy transaction processor
 - Specification for transfer tool interactions
 - Specification/process for transfer area management
 - Description of processes, including state definitions
 - Estimate of task size and complexity for software and process changes
 - Identify recovery scenarios
 - Describe replication selection
 - How is a Peer chosen?
 - How is the service model encoded?

- **Services & Policies Committee (Advisory Committee?)**

- Work with the Services manager to plan and develop the future course of service offerings through DPN, as well as serving key insight into the membership and the needs of the preservation community at large.
- Pilot Tasks
 - Evaluate/recommend model for 1 copy in DPN
 - Does this require registration of Peer-local copies?
 - Evaluate/Recommend model for 2 copies in DPN
 - What is the length of service
 - 5, 7, 10, 20 years?
 - How do the models assure recoverability?
 - Service descriptions for each model
 - Differentiate offering from a disaster recovery service
 - Does this overlap with partner offerings?

Policy & Procedural Considerations

1. Pilot Specific considerations
 - a. Pre-pilot
 - i. # of participants
 - ii. Pilot agreement document
 1. Identify content to be transferred
 2. Protocols for transfer/tools
 3. Data retention after the pilot?
 - iii. Set criteria for success
 1. Content successfully transferred
 2. Content moved to Peer secure storage
 3. Content registered in DPN
 - iv. Contractual issues
 1. Peer deposit agreement with DPN
 2. Storage agreement/SLA with DPN

3. Hosting DPN software component

b. Stage 1

- i. Workflows and transaction definition
 1. Each peer identifies a set of content to be deposited
- ii. Share networking information/open firewalls
- iii. Configure transfer mechanisms
- iv. Interface with transfer tools
- v. Initial transfer of content - proof of concept

c. Stage 2

- i. Hosting DPN software component - requirements/constraints
- ii. Automation/scripting of content transfer
- iii. Generate test bags & validate
- iv. DPN transaction and metadata exchange
 1. Register a bag via API services
 2. Create replication entries
- v. Peer node retrieves replication requests
- vi. Peer node pulls data via the transfer mechanism
- vii. Review of data security standards

d. Post Pilot

- i. Peer fixity calculation & reporting
- ii. Storage equity calculation
- iii. Determining transaction costs
- iv. Identify software and process changes needed
- v. Identify peer node technical staff required
- vi. Considerations for optimizing transfers
- vii. Validate process/ability to recover content
- viii. Recommendations for transitioning from a pilot to a production service

2. Operational & Contractual considerations

- a. Legal model for peer long term storage allocation to DPN
- b. Demonstrating transactional viability for
 - i. Data movement
 - ii. Content registration & tracking in DPN
- c. Will some data be stored using the traditional DPN service model?
- d. What is the exit strategy for peer arrangements?

Prerequisites

1. Identification of at least two participants
2. Outline of legal and procedural agreements
3. Basic content selection guidelines
4. Timeframe for pilot activities

Glossary & FAQ

- http://bit.ly/DPN_Glossary
- <https://groups.google.com/forum/#!forum/dpn-faq>

