Requirements and Assumptions

DPN requirements and assumptions

In the development of the system, legal and technical constraints have been imposed that change some assumptions, as originally conceived.

Those changes do not affect the fundamental assumptions of the DPN concept.

DPN is dark, geographically and technologically diverse replication, with individual and centralized audit and succession.

"The intent of DPN remains the same. Build a system that ensures that deposited material remains available to future generations by architecting around single points of failure broadly conceived (i.e., technical failure, political failure, organizational failure, geographic failure)."

- DPN is a federated preservation network of independent preservation archives
    - All archives act to preserve content, not just as remote storage.
- DPN has agreements to allow for Succession of content in the event an archive can no longer perform its function as an archive.
    - This is currently envisioned as a “Quit Claim” framework
    - Based on input from the legal team, comprised from the Node institutions, we will update recommendations
    - The successor of content will take on the responsibilities of and become the Administrative Node for the failed archive

- DPN, as much as possible, has independent preservation implementations at each node.
- DPN communications follow best practices and use mutual authentication over secure channels.
- DPN provides for replication of content from a Ingest Node to Replicating Nodes
    - DPN provides recovery of lost content at any Node
    - DPN core service keeps 3 copies of ingested objects
- DPN provides a service model so that any of the Nodes can ascertain if their content is valid.
- DPN enables centralized auditing
    - Process audit at the DPN federation
    - Process audit at the Node
    - Audit trails for events
    - DPN supports Content Fixity audit
- DPN provides status reporting, activity, logging, traffic, volume, events, etc.
    - Global (across federation)
    - At the Node level
    - Periodic Reporting
    - Significant event reporting, for example succession events, loss of content, etc.
- DPN supports
    - a DPN UUID (globally unique identifier for each bag deposited)
    - a common lightweight wrapper (bag) for content transfer
    - retention of ingested content indefinitely
        - Content may be de-accessioned upon extenuating circumstances (e.g. court order)
    - duplication of critical metadata in the 'registry' and also in the content bags
    - durable and persistent communication methods to support unreliable networks and node failure
    - a distributed model, assuming eventual consistency (CAP Theorem) for replication of content and registries

- DPN content and services are distributed and federated.
- DPN supports succession and brightening of content.
- Implementations are de-coupled implementations and architecturally distinct, as practicable, but the communication methodology is shared, resilient, and redundant.
- DPN objects are preserved by all Nodes and support DPN preservation functions.
- DPN has decoupled inter-node communication channels for
    - content transfer
    - process control.
- Communication between nodes is not dependent upon other nodes.

Production Deliverables - Jan 2016

- There is no expectation that we’ll have signed SLA’s before launch
    - DPN will have agreements to allow for Succession of content in the event an archive can no longer perform its function as an archive.
- SLA’s with Ingest Nodes/Administrative Node
- SLA documents between Ingest Nodes & Depositor shared and with legal staff
    - We will require in the SLA that one of the OTHER Administrative Nodes will become the Administrative node for the failed Node.
        - Act as the restorer of content for clients of the former administrative node
        - In the event of a Succession occurrence, all replicating nodes will recognize the new successor and act in accordance with prior agreements held by the former archive
    - Replication and update will take time - We are aware of this, might need to put something in the SLA recognizing this fact
    - SLA will state that: The communication layer is shared, but the repository layer is not
- Depositor will be able to give us stuff and we can put it into storage
- There will be no expectation of global (DPN Level) fixity checking at launch
    - Initial fixity checking will occur at ingest
    - Each Node will check fixity according to their local policy
- Some cursory reporting available to Depositor

    - Notification of ingest and replication will be provided by the Ingest Node to the depositor
- Replication of content from the Administrative Node to all Replicating Nodes
    - Make sure we know how many Replicating Nodes will be storing the content
    - Make sure we know when the Ingest Node will store the ingested content and when they won't
    - Make sure we have a clear idea of what kind of storage is available at each
- The Ability to recover content to the depositor will be supported via the Ingest/Administrative node
- Maintain a registry for objects, create transfer records for replicating nodes, update status of an existing object
    - Track stored status of replicated transfer
- Agreement on the bag size limits that can be replicated across the Nodes
    - 250 Gig bags are the uppermost limit for this release
- Bags will be validated upon receipt by the Ingesting Node and the Replicating Node
    - Validation means:
        - We will validate all of the files are present and the checksums match the manifest
        - The structure of the bag will be validated according to the DPN specs
- In-Person Post-Mortem/Planning for Phase II - July 16th & 17th

6-Month Post-Launch Roadmap Deliverables (Jun 2016)

- Clear idea of what kind of storage capacity is available at each Replicating Node with a framework for deciding to which nodes objects are deposited (Internal Documentation)
- Auditing consistency of registry
- Auditing the local storage inventory of registry
- Fixity tracking - when replicating nodes are doing auditing - including node that performed the fixity check (as part of the provenance/history of the bag)
- Support multiple fixity types across the federation
- Ongoing fixity checks by each node with reporting out to DPN administration

12-Month Post-Launch Roadmap Deliverables (2017)

- Depositor Dashboard
    - DPN will provide status reporting, activity, logging, traffic, volume, events, etc.
        - Global (across federation)
        - At Node level
        - Periodic Reporting
        - Significant event reporting, for example succession events, loss of content, etc.
    - Billing information
- Bag Discovery & Retrieval Request Mechanism
    - a depositor wants all of their bags will satisfy some criteria
    - Determination and tracking of depositor assets