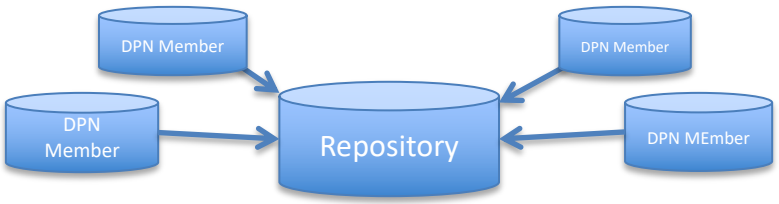
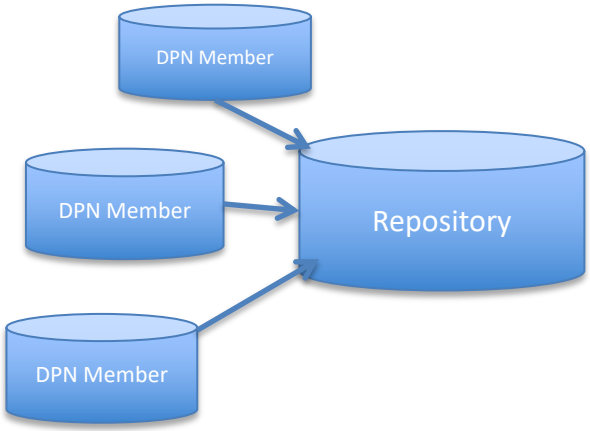
A Report On DPN's Emerging Architecture, System Protocol, and Service model

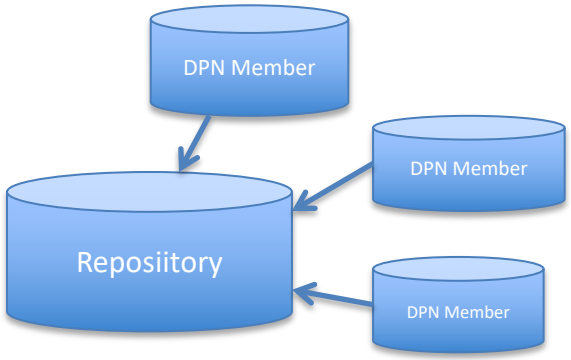
CNI Spring Meeting
April 5, 2013 * Austin, TX

Tom Cramer, James Simon

What Is DPN?

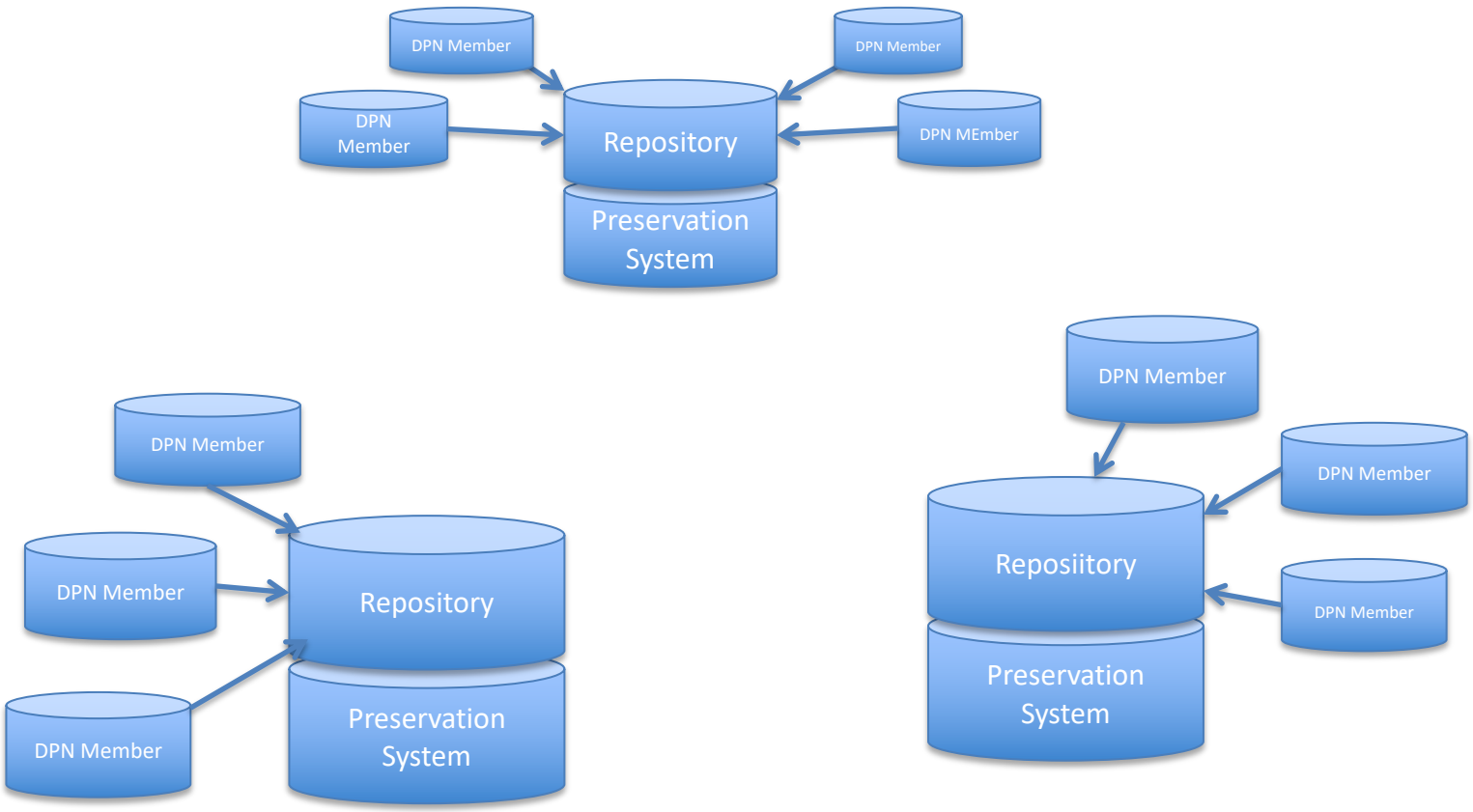






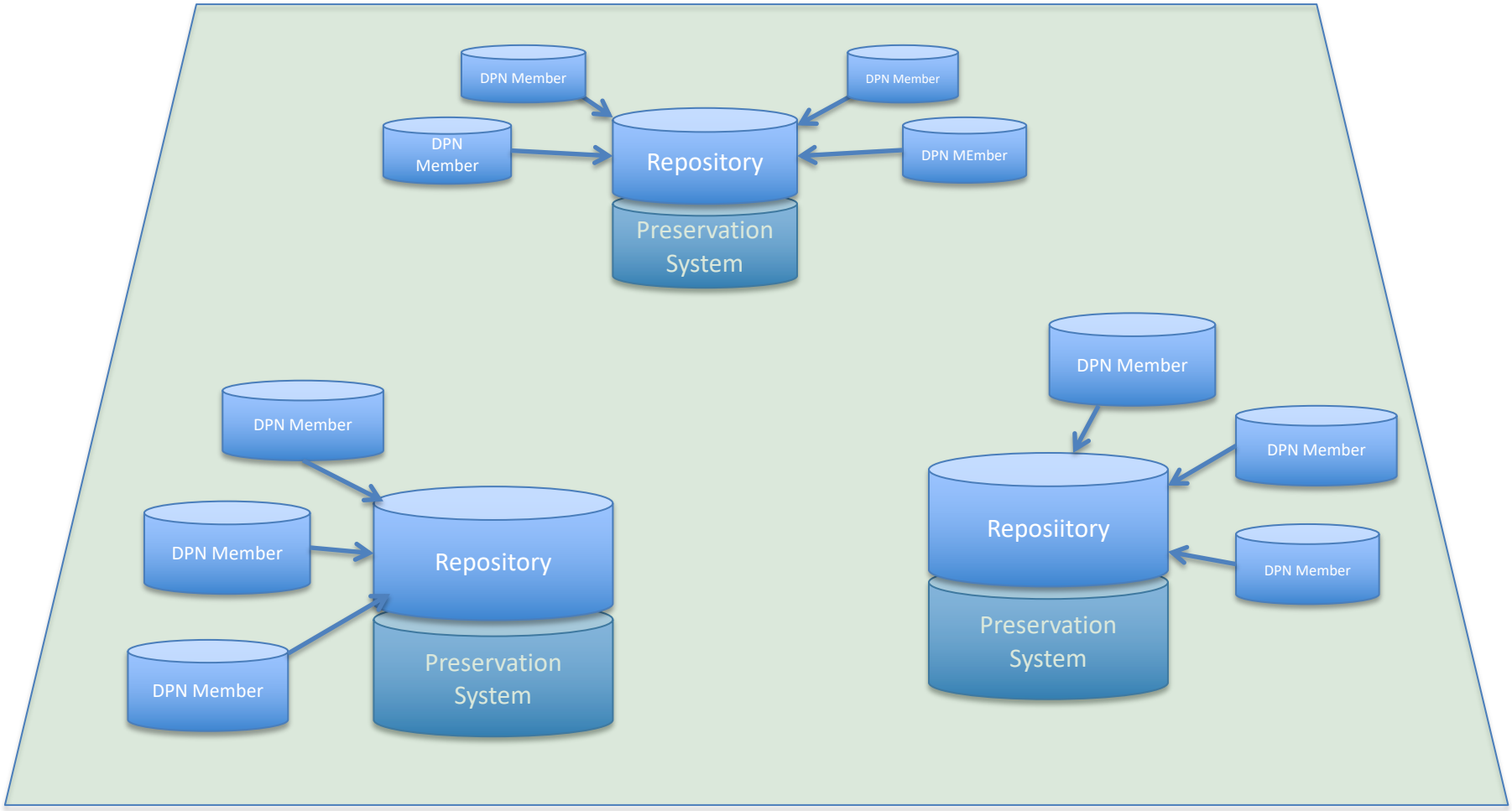
57 member organizations cooperatively investing in long-term, scalable, digital preservation

What Is DPN?



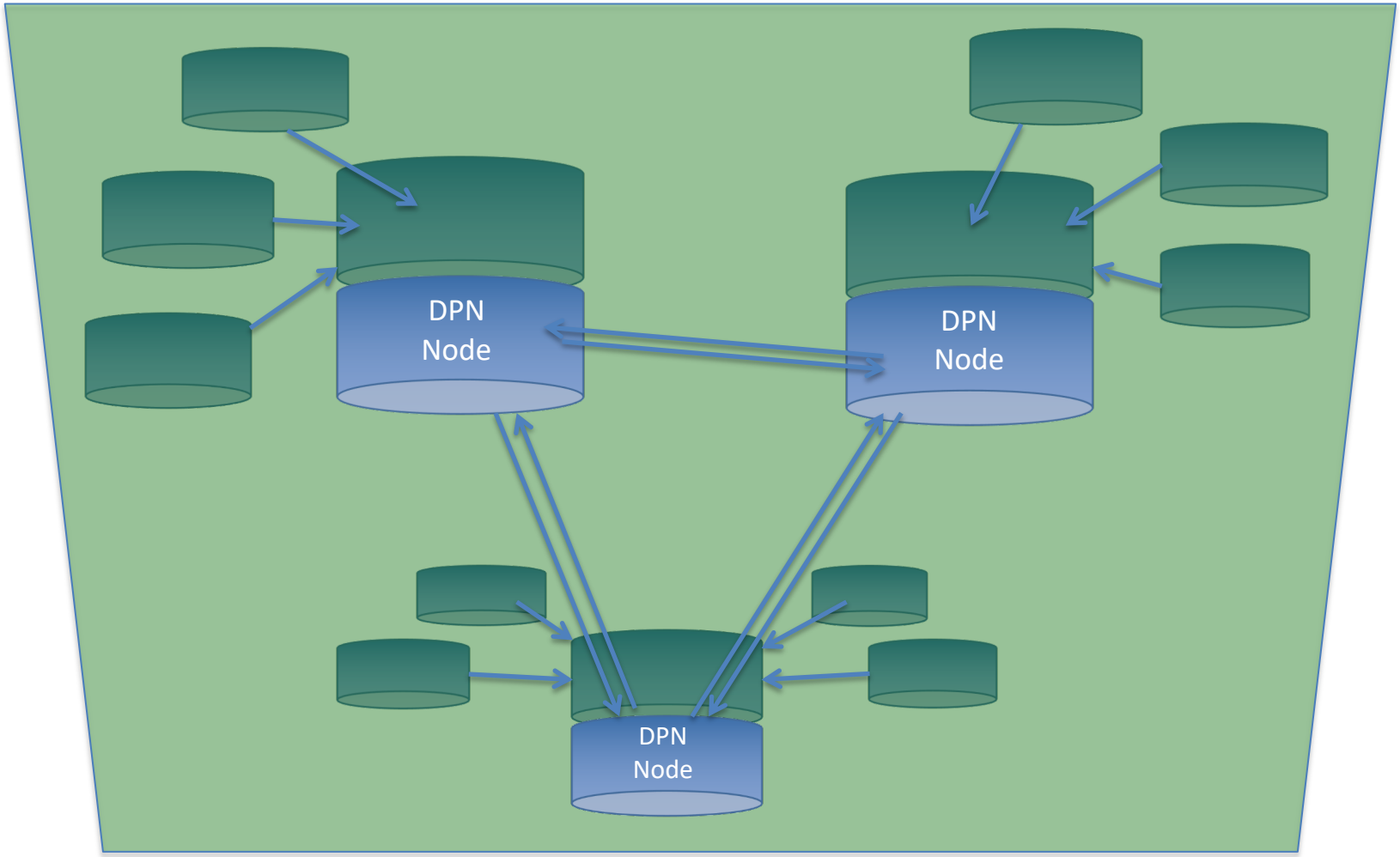
technical staff and systems from 5 large scale preservation repositories

What Is DPN?



…working groups of experts in succession rights, business services, communications and research data…

What is DPN?



All building a digital preservation backbone for the academy

Initial DPN technical partners

Initial DPN launch will feature five nodes:

- Academic Preservation Trust (APTrust)
- Chronopolis
- HathiTrust
- Stanford Digital Repository (SDR)
- University of Texas Data Repository (UTDR)

And a participating partner:

- DuraSpace

The DPN Technical Team

APTrust

- Scott Turnbull
- Tim Sigmon
- Adam Soroka

Chronopolis

- David Minor
- Mike Smorul
- Don Sutton

DuraSpace

- Andrew Woods

HathiTrust

- Sebastien Korner
- Bryan Hockey

Stanford

- Tom Cramer
- James Simon

Texas Data Repository

- Ladd Hanson
- Christopher Jordan

What Does DPN Do?

1. Establishes a network of heterogeneous, interoperable, trustworthy, preservation repositories (Nodes)
2. Replicates content across the network, to multiple nodes
3. Enables restoration of preserved content to any node in the event of data loss, corruption or disaster
4. Ensures the ongoing preservation of digital information from depositors in the event of dissolution or divestment of depositors or an individual repository
5. Provides the option to (technically and legally) "brighten content" preserved in the network over time

DPN Benefits

1. Resilience
2. Succession
3. Economies of scale
4. Efficiency
5. Extensibility
6. Security

Critical Assumptions & Definitions

- All content enters DPN by deposit into one of the DPN Nodes.
- The “First Node” is the point of entry for a given piece of content; Nodes with copies of this content are “Replicating Nodes”.
- DPN Members will work directly with an individual DPN Node to negotiate contracts and determine service levels
- Service levels and contracts will reflect “standard” DPN services; they may also reflect the First Node’s unique offerings in terms of access, hosting or other services.
- Content in Replicating Nodes will be held “dark”, and inaccessible except for preservation actions.

Critical Assumptions & Definitions

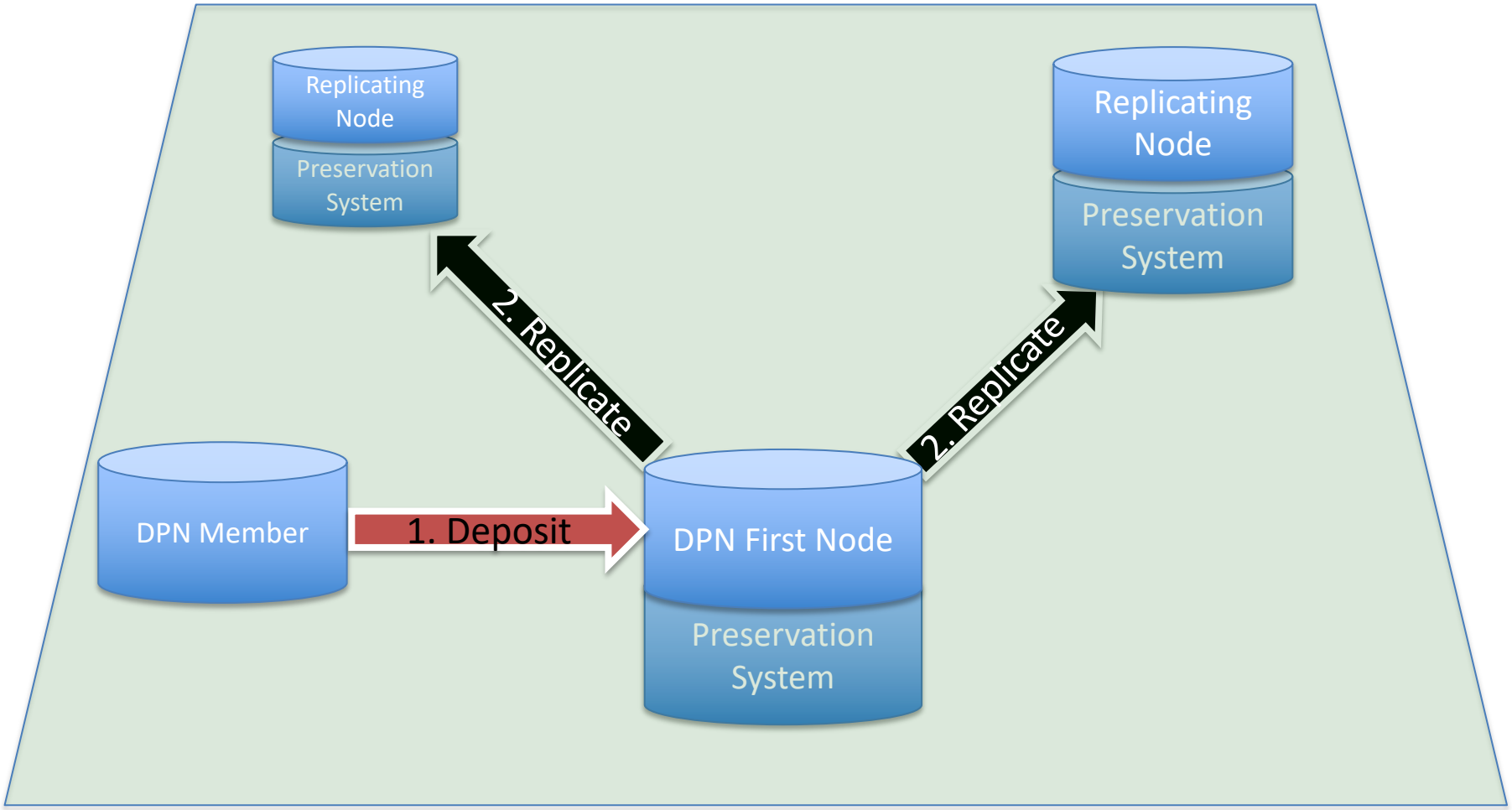
- DPN shall redistribute preserved content as Nodes enter and leave the Network, ensuring continuity of preservation services over time.
- DPN will provide a large-scale network of dark archives that enable the opportunity to brighten content in the future, but does not mandate how this is done.
- Depositors, First Nodes and their designated communities will collaborate to ensure that the information contents of DPN deposits are accessible for reuse in the future, using the appropriate (and evolving) community standards for any given set of content.

Specifications (sample of…)

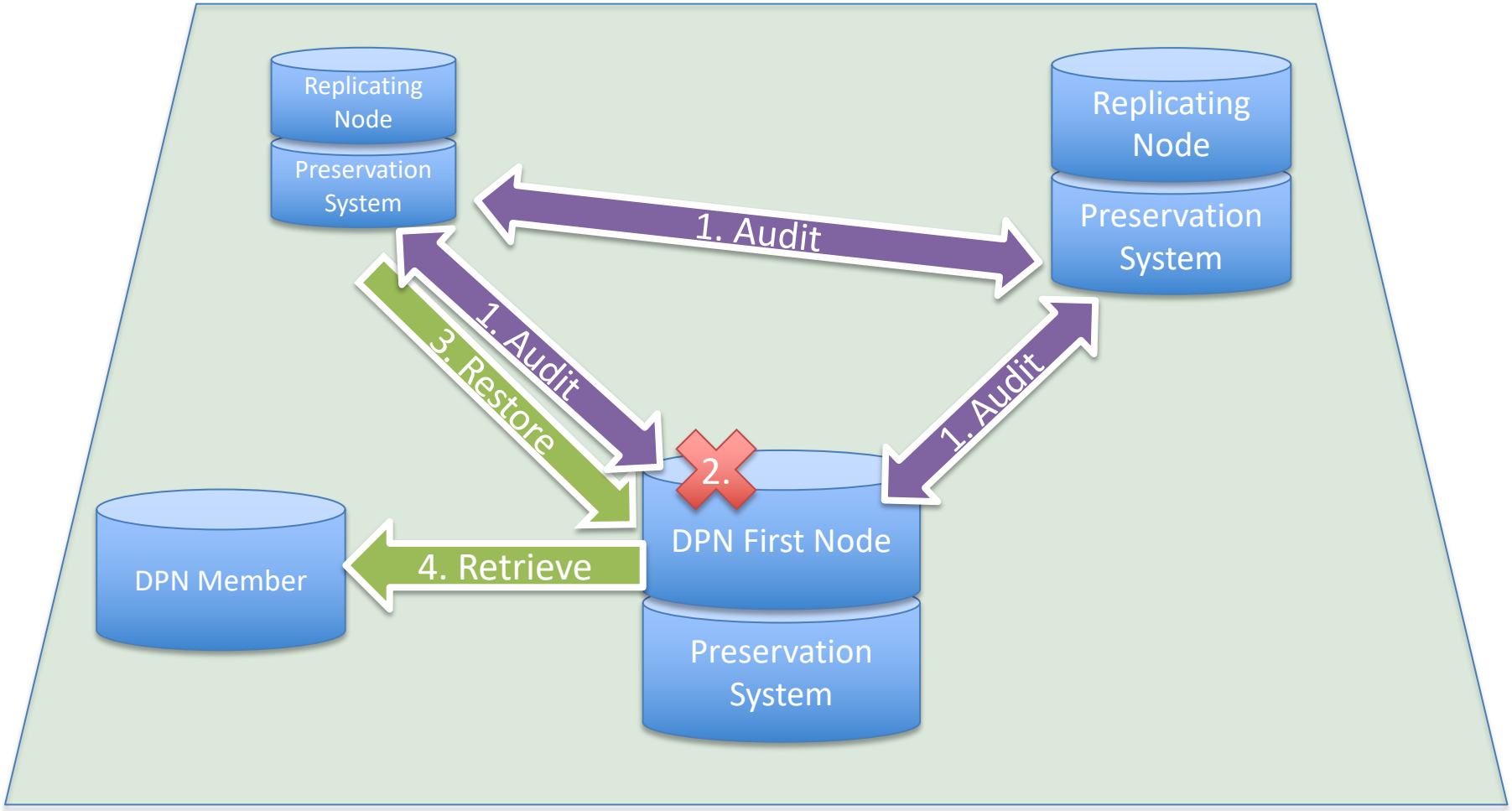
1. *DPN will make multiple copies of content from a First Node to Replicating Nodes.*
5. *DPN will repair or replace any replicated content at any node when corruption is detected.*
7. *DPN will assure the security of replicated content during transmission so that no content is lost, corrupted, or exposed.*
16. *DPN will be able to support the introduction or exit / cessation of DPN Nodes by redistributing content among new/continuing nodes to ensure sufficient copies are kept according to policies.*

See https://wiki.duraspace.org/display/DPNC/Specifications

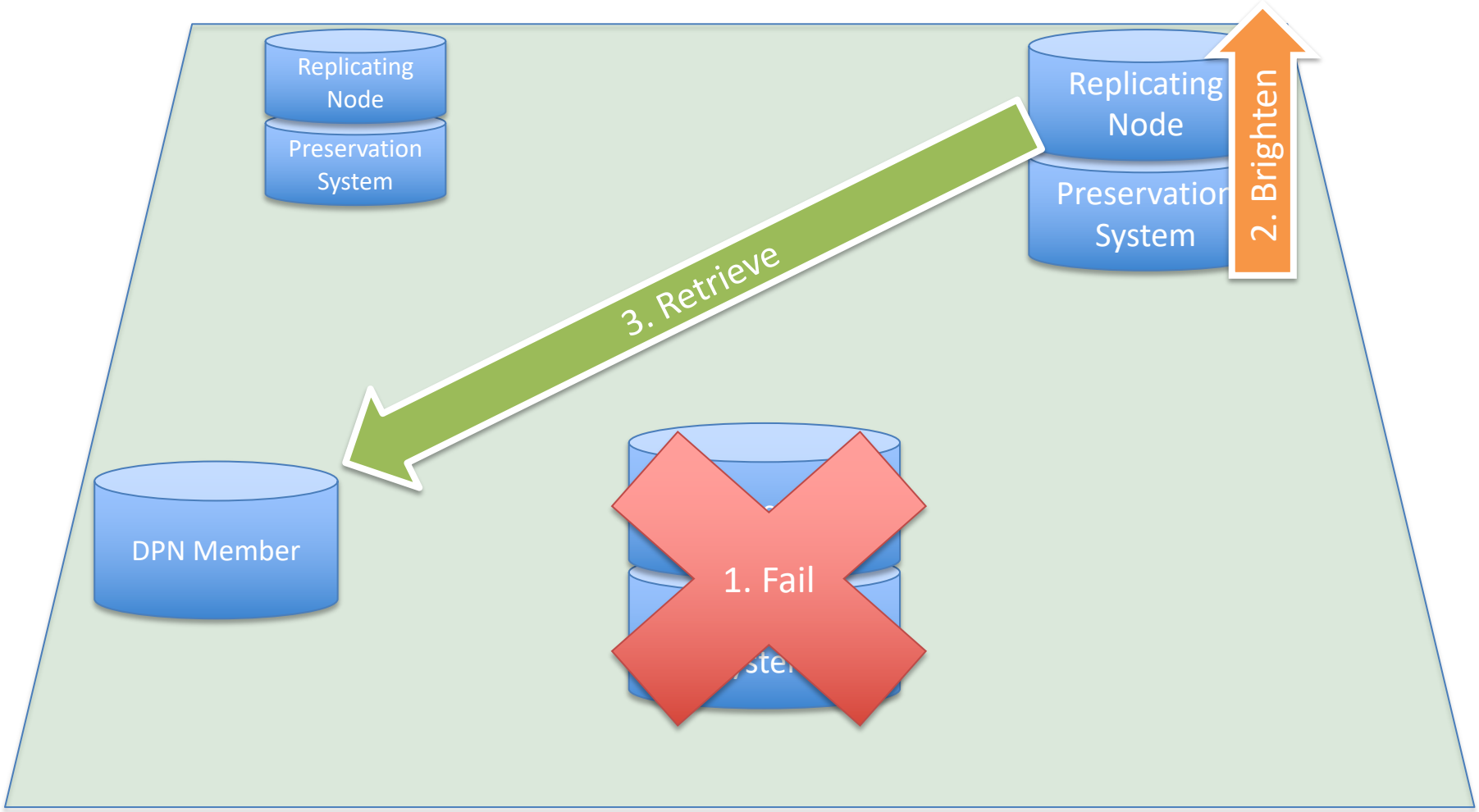
Scenario 1: Ingest & Replication



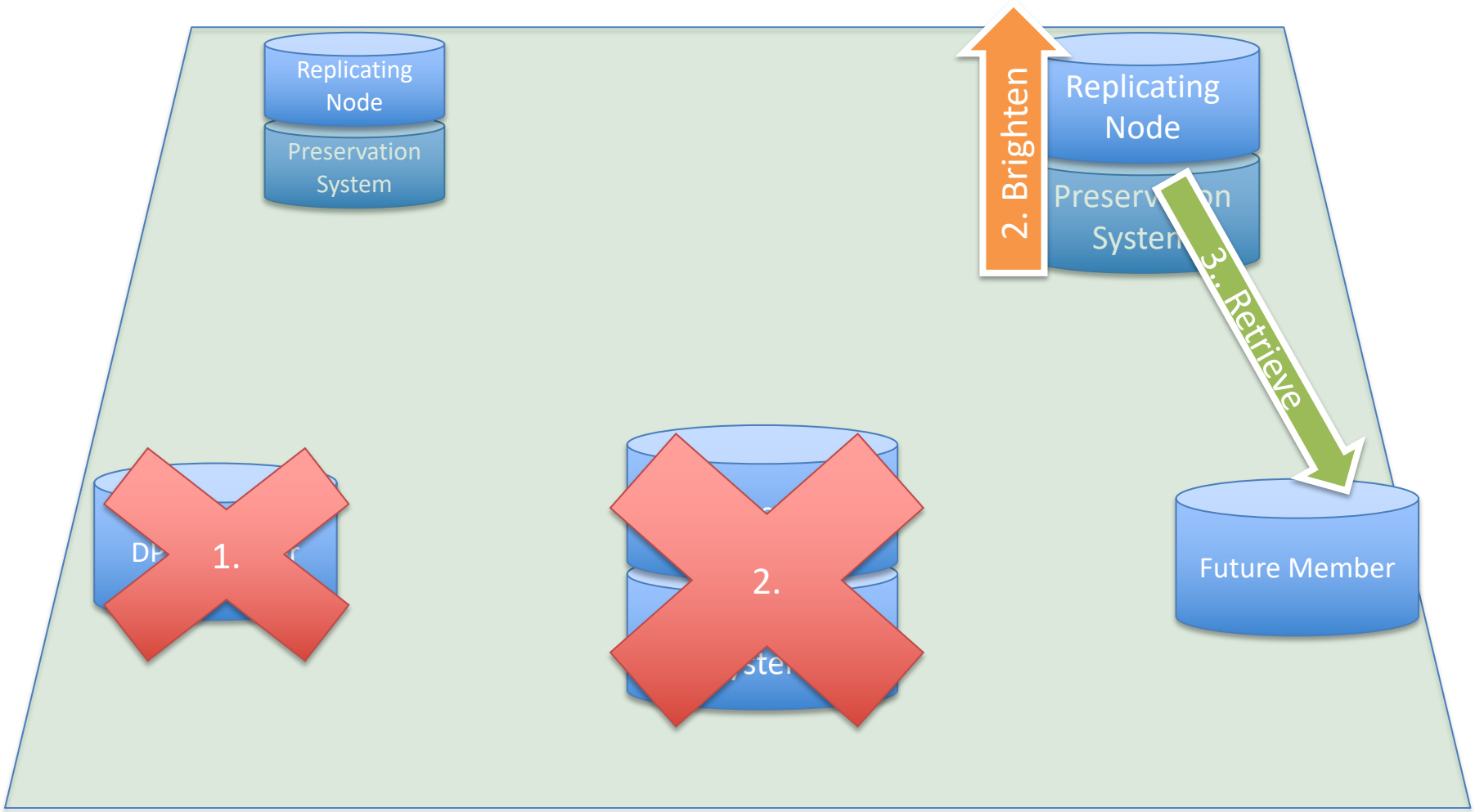
Scenario 2: Restoration of Content



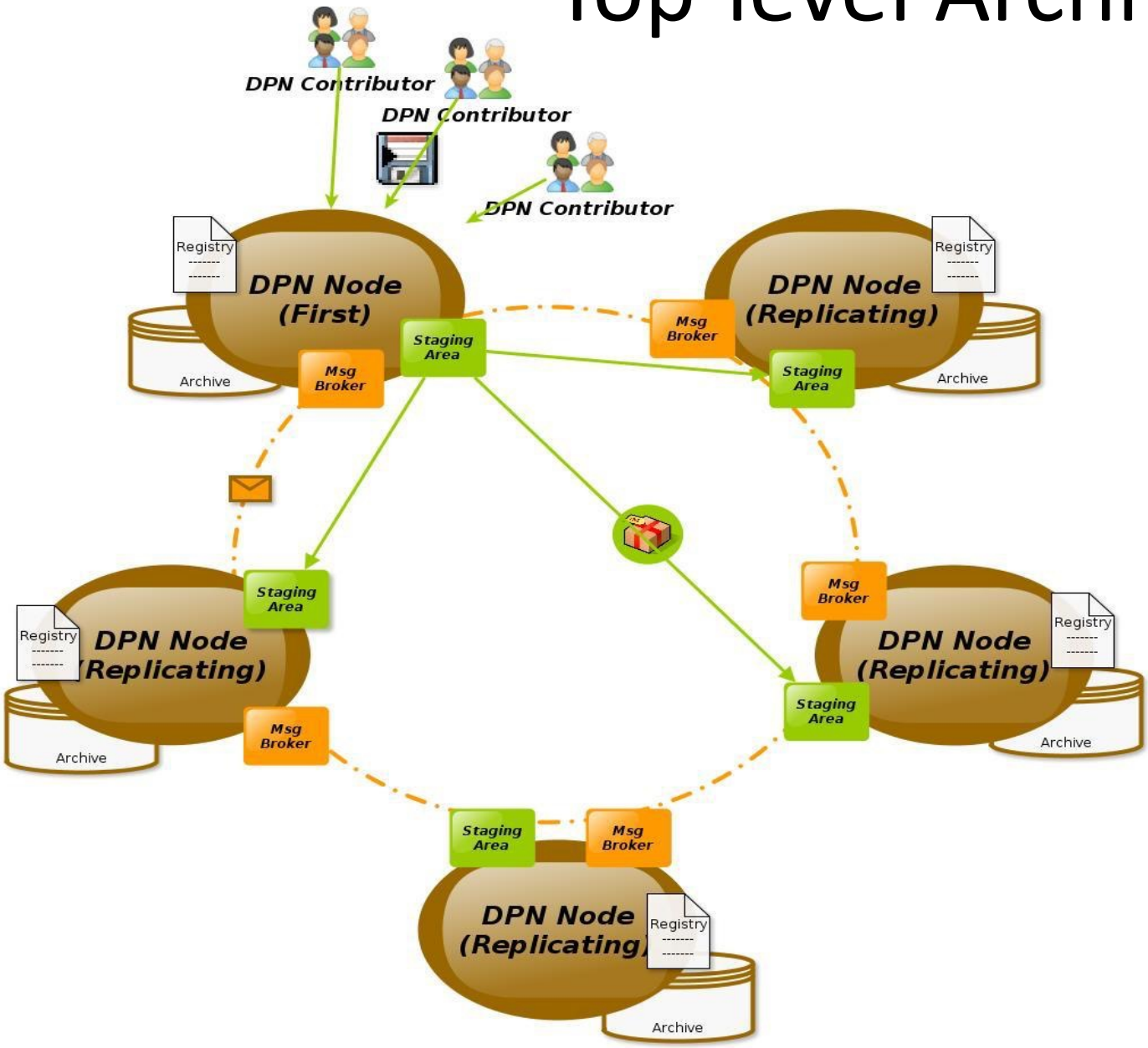
Scenario 3: First Node Cessation



Scenario 4: Successioning

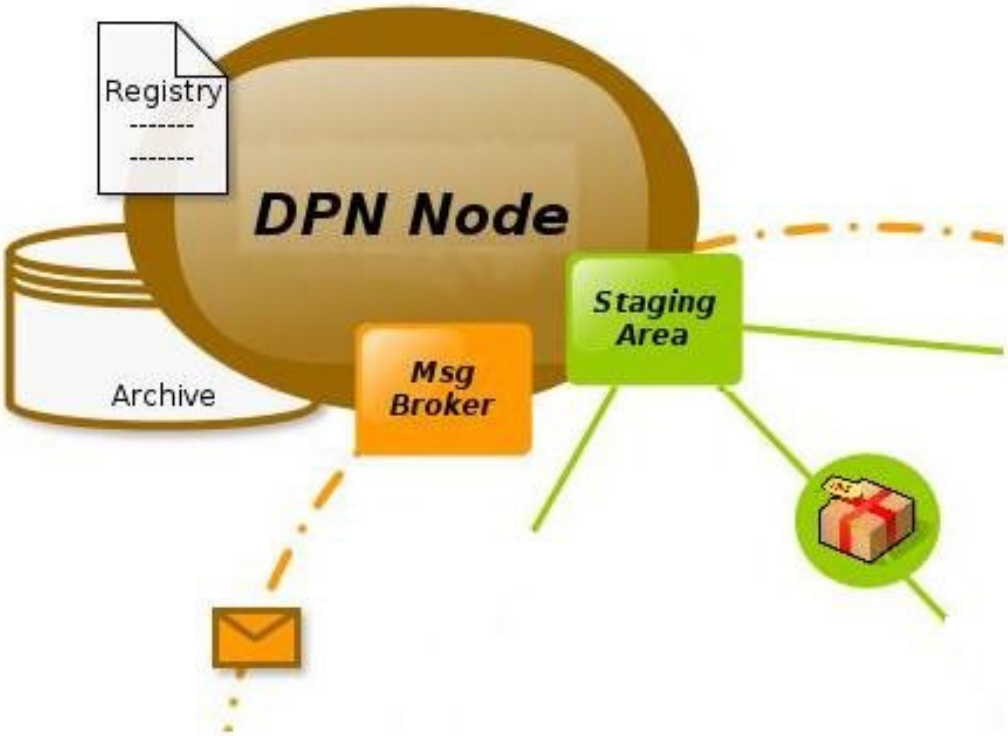


Top-level Architecture



Infrastructure Components

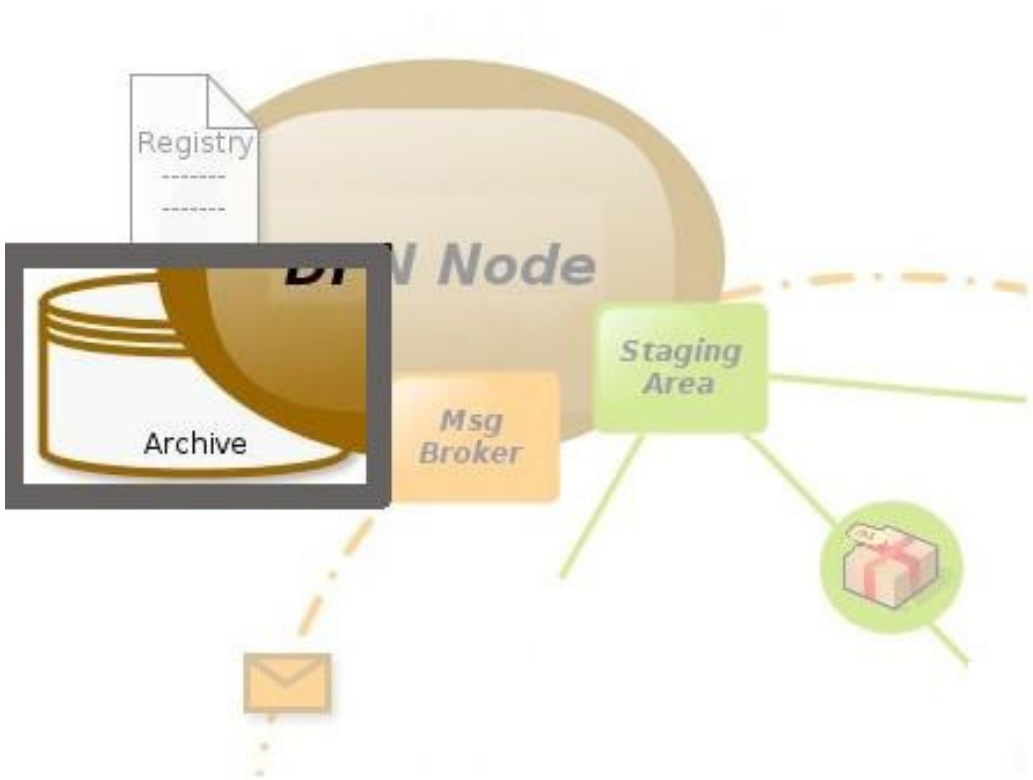
- *Institutional Archive/Repository*
- *Federated Messaging*
- *Distributed Registry*
- *Transfer Mechanisms*
- *Content Packaging*
- *Security and Encryption*



Infrastructure Components

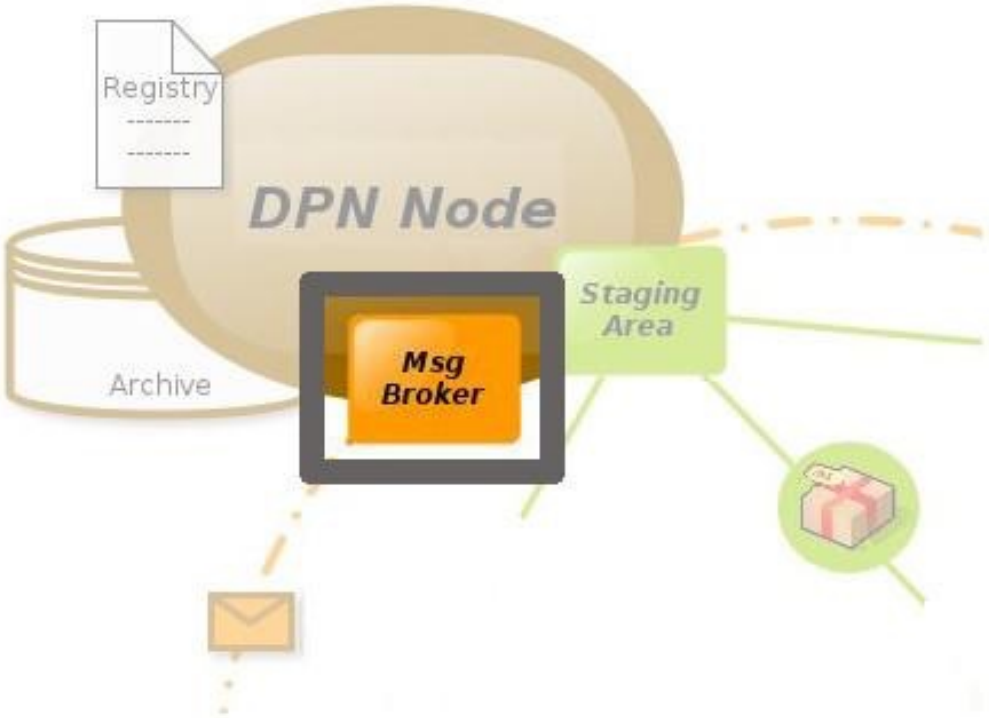
- ***Institutional Archive/Repository***

- *Federated Messaging*
- *Distributed Registry*
- *Transfer Mechanisms*
- *Content Packaging*
- *Security and Encryption*



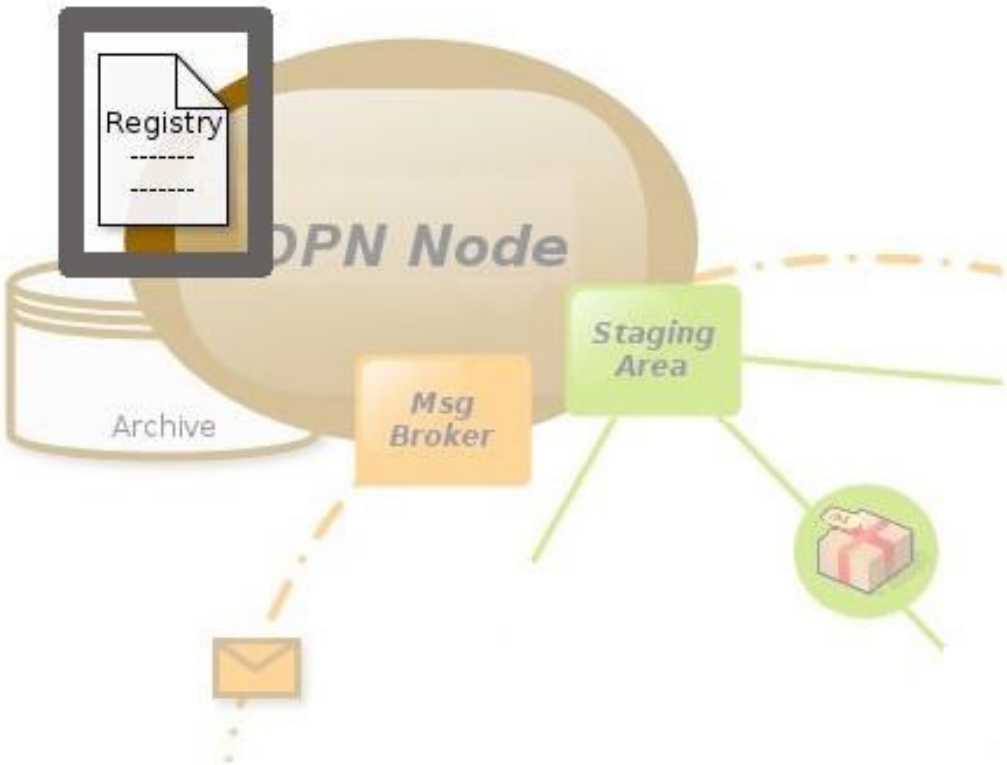
Infrastructure Components

- *Institutional Archive/Repository*
- ***Federated Messaging***
- *Distributed Registry*
- *Transfer Mechanisms*
- *Content Packaging*
- *Security and Encryption*



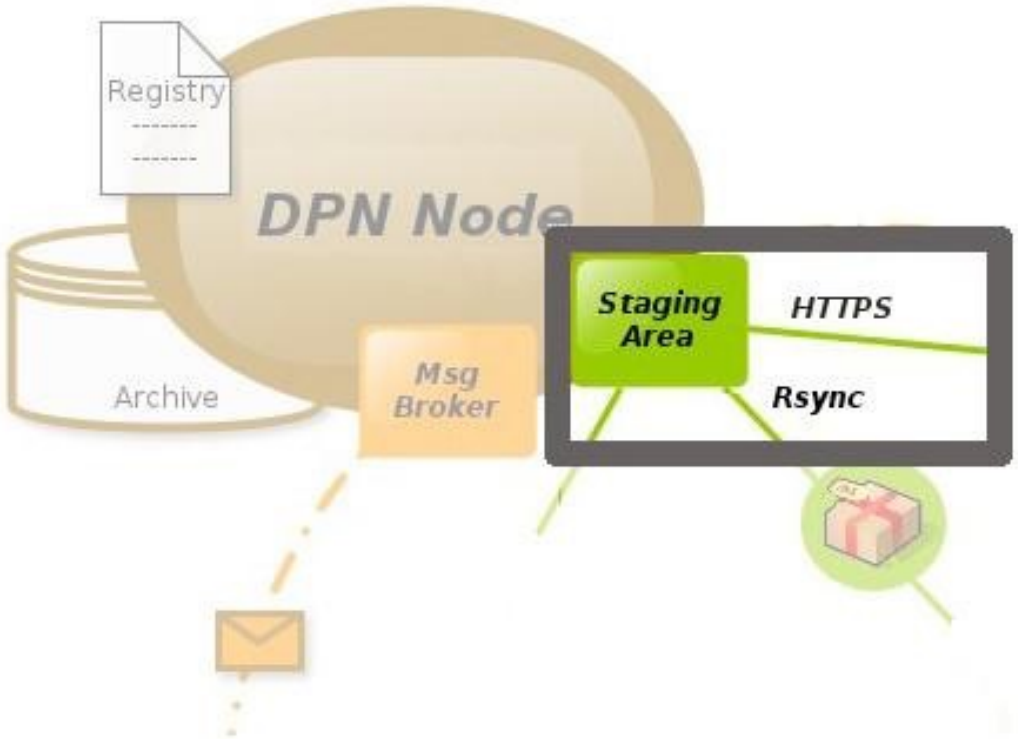
Infrastructure Components

- *Institutional Archive/Repository*
- *Federated Messaging*
- ***Distributed Registry***
- *Transfer Mechanisms*
- *Content Packaging*
- *Security and Encryption*



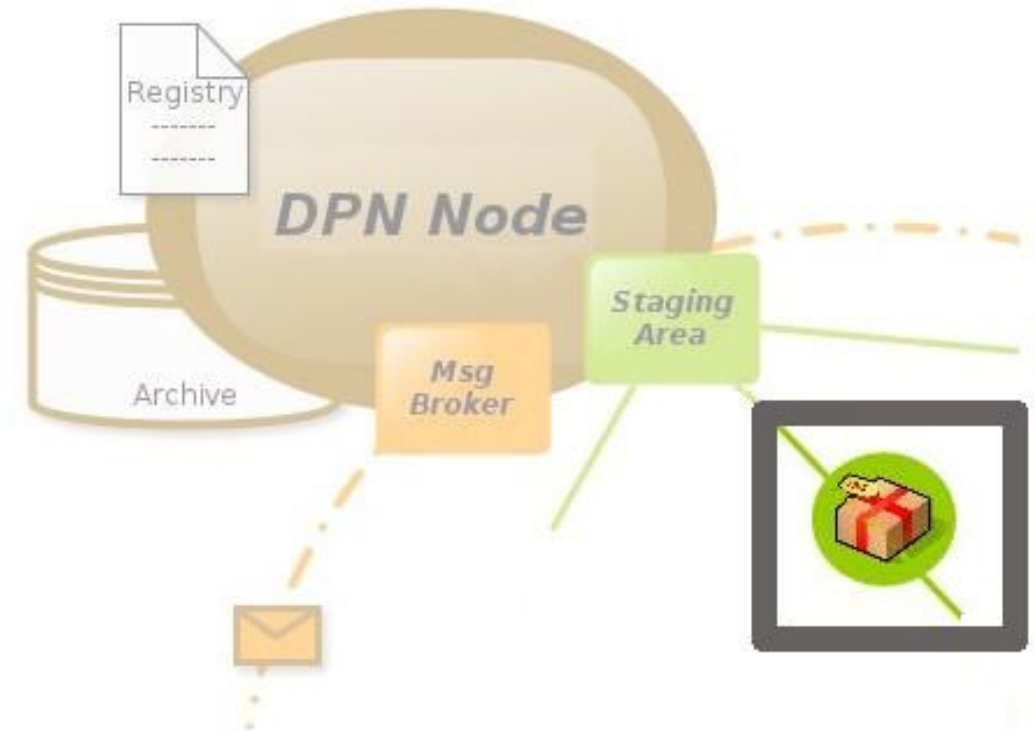
Infrastructure Components

- *Institutional Archive/Repository*
- *Federated Messaging*
- *Distributed Registry*
- ***Transfer Mechanisms***
- *Content Packaging*
- *Security and Encryption*



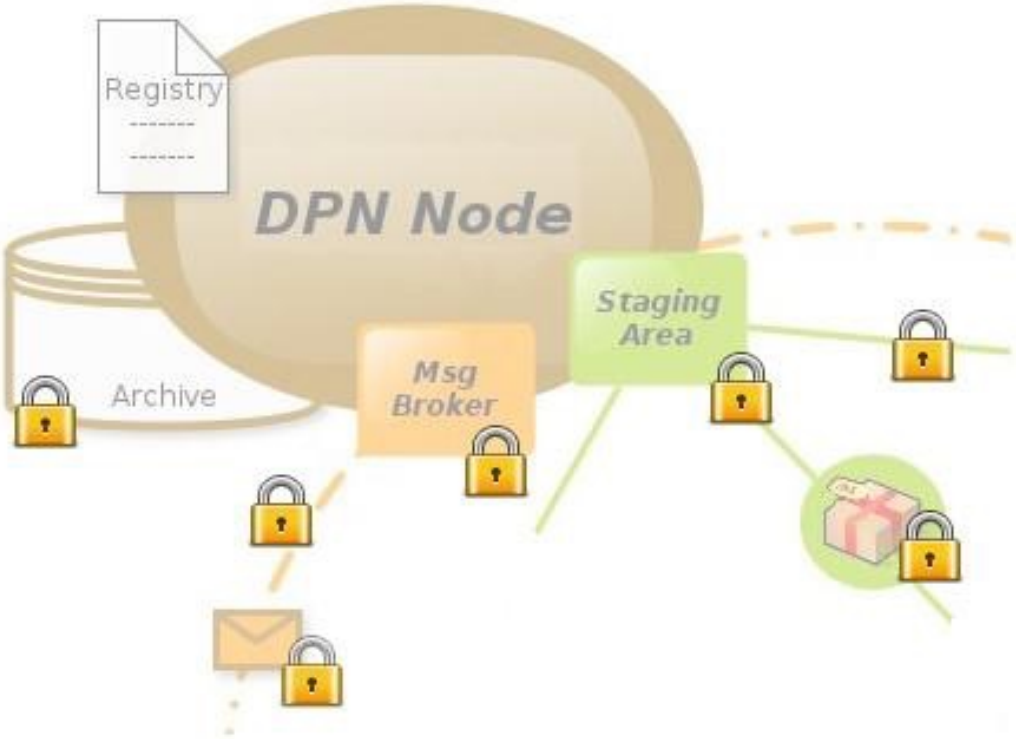
Infrastructure Components

- *Institutional Archive/Repository*
- *Federated Messaging*
- *Distributed Registry*
- *Transfer Mechanisms*
- ***Content Packaging***
- *Security and Encryption*



Infrastructure Components

- *Institutional Archive/Repository*
- *Federated Messaging*
- *Distributed Registry*
- *Transfer Mechanisms*
- *Content Packaging*
- ***Security and Encryption***

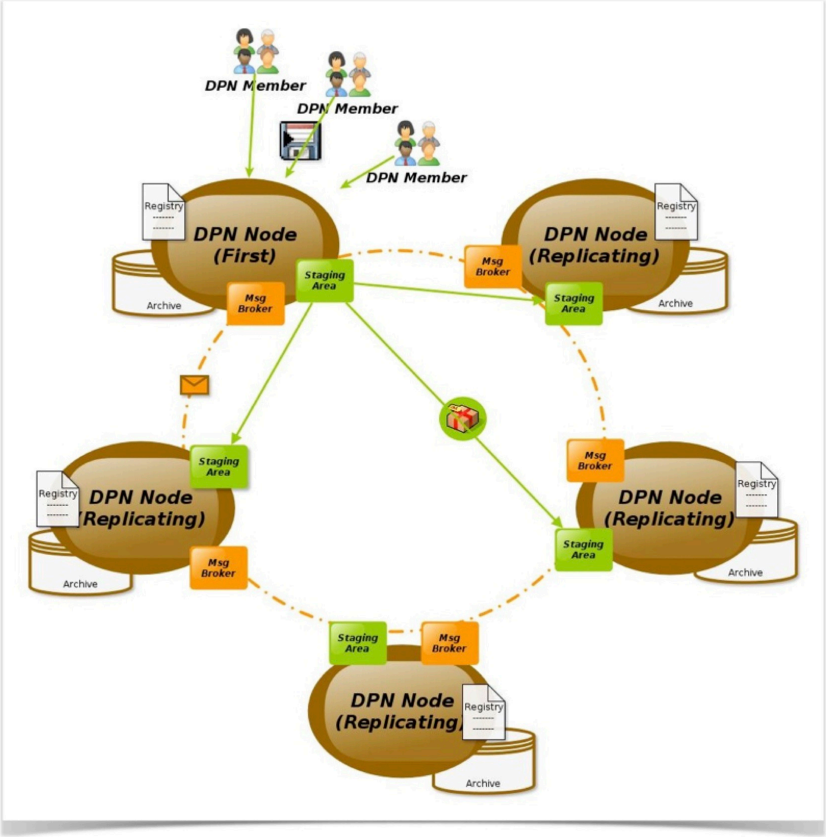


Architectural Overview

- Architectural Premise
  - *Core capabilities founded on proven institutions and repositories*
- Design Considerations
  - *Distributed Nodes, loosely coupled*
  - *Standards and protocol-based integrations*
  - *Separate implementations*
  - *Distributed infrastructure*

The DPN Federation

- Each node in the federation acts as both a First Node, and a Replicating Node
- The Nodes will put content into DPN from their own repository, or from a DPN member
- There is a new hierarchy for content, with DPN First Nodes having responsibly for preservation activities within the DPN federation
- Part of those activities include keeping agreements up to date, transitioning fixity to new algorithms, fixity audit across the federation, logging, etc.



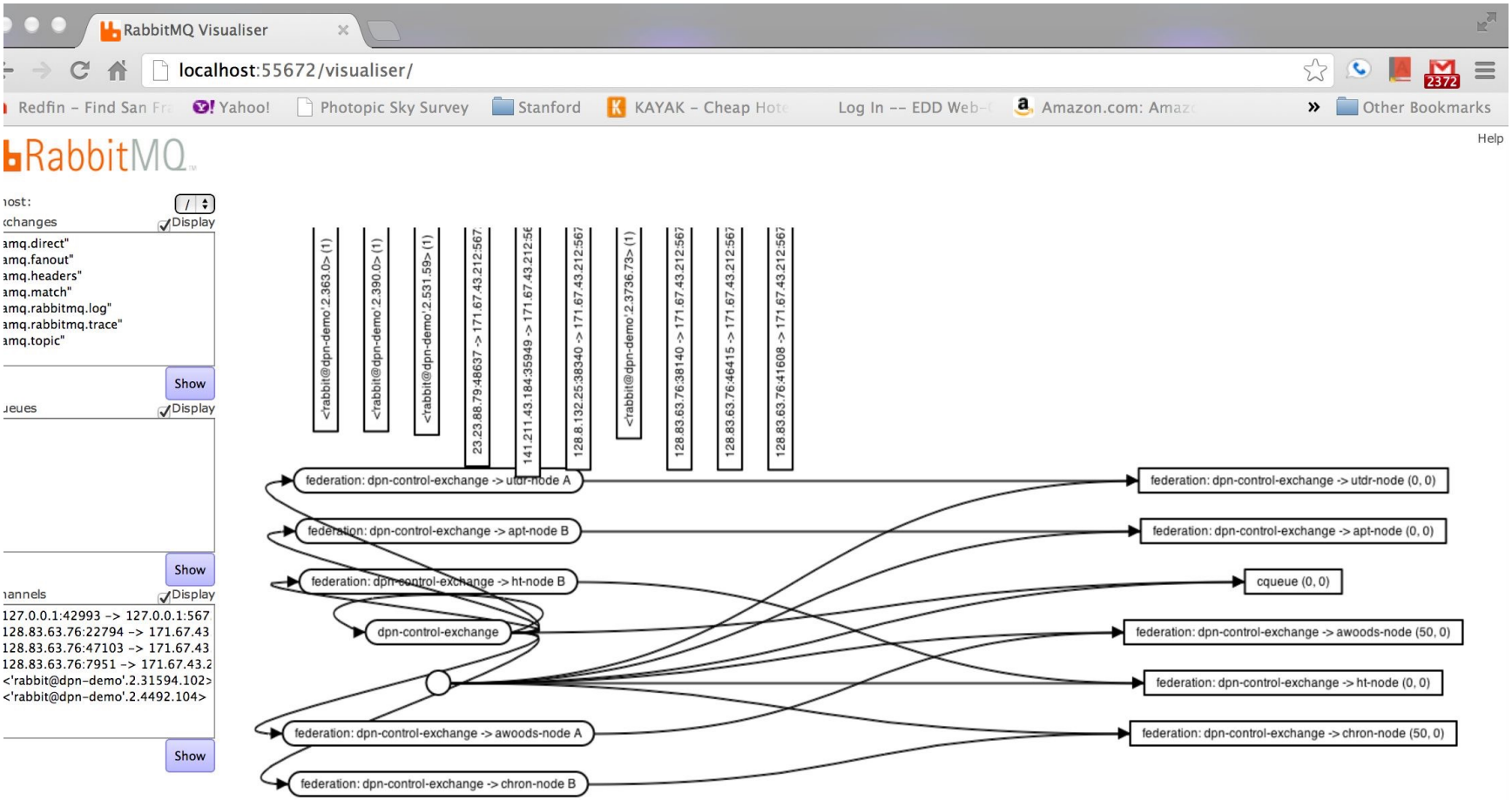
Components in Technical Architecture

- Messaging infrastructure to support federated services
- Registry to track objects within the federation, including copies, version, rights, brightening information
- Transfer mechanisms (rsync, https, gridFTP, etc.)
- Private PKI for securing transport layers
- Logging and reporting
- Other components we implement separately, but may be common, for example a secure transfer area.
- DPN objects that hold administrative content such as DPN framework agreements, DPN bagit profiles, versioned Brightening information for a collection/repository

Messaging

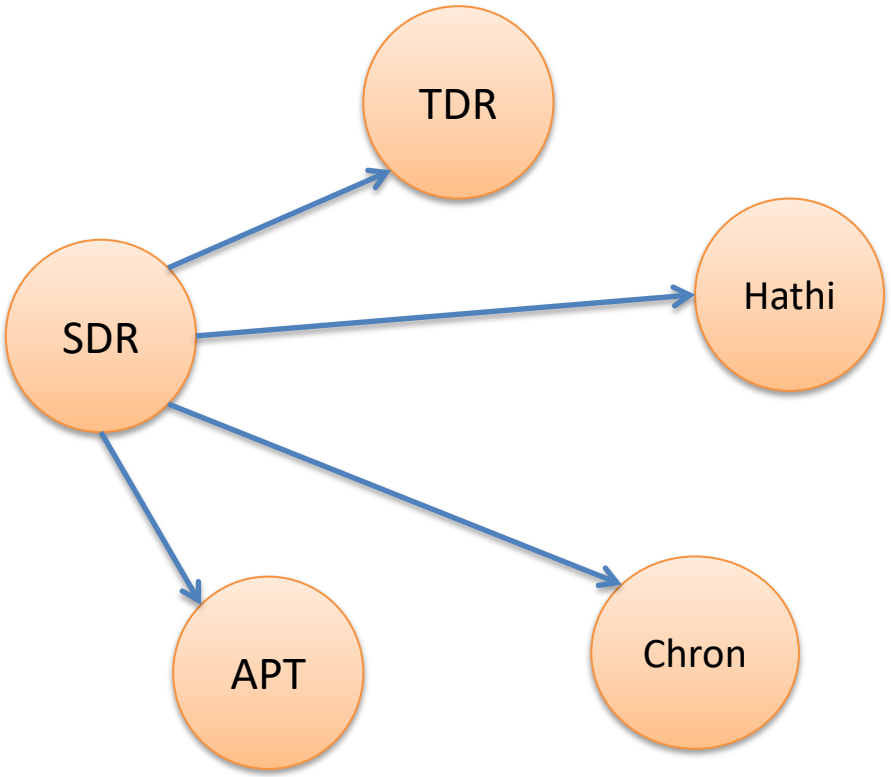
- DPN uses messaging for in band communication and control and replication and services are handled out of band
- Using RabbitMQ message brokers, which support AMQP (Advanced Messaging Queueing Protocol)
- RabbitMQ also supports federation easily
- DPN messaging model uses Topic Queues for broadcast messages and direct queues for one-to-one communication between nodes

Federation View



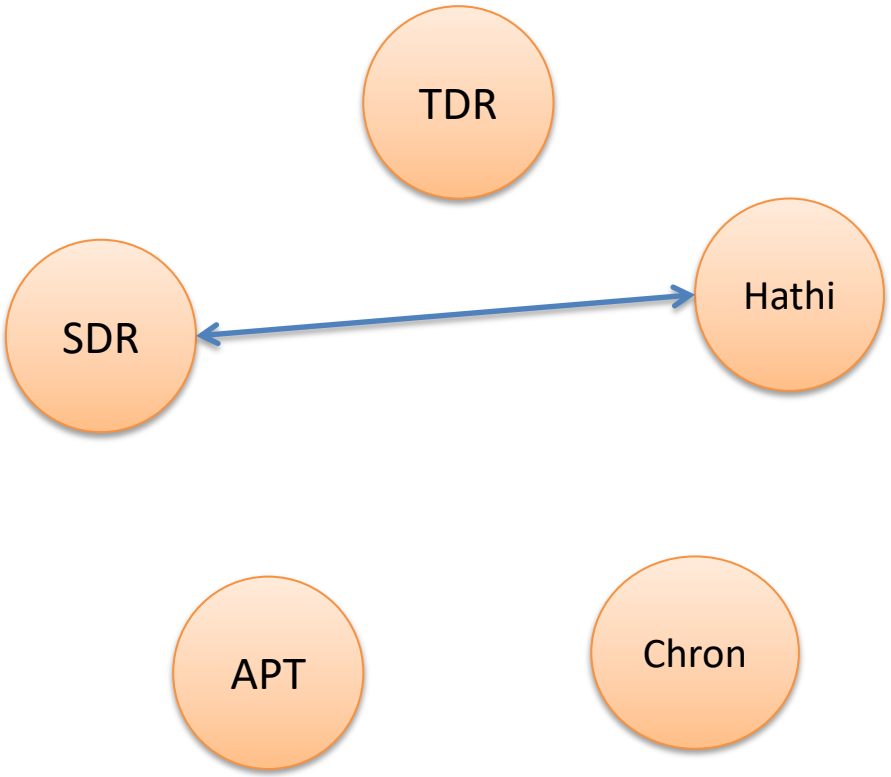
Messaging Model

- Broadcast messages are sent to all nodes at the same time
- Each broker is federated to all of the others, so if one broker is down it is still possible to communicate



Messaging Model

- Direct messages are between two nodes, used for replies in a message sequence
- Broker federation still applies, so communication channels are redundant



Messaging Control Flows

- Message control flows are transactional, and asynchronous

Examples of control flows

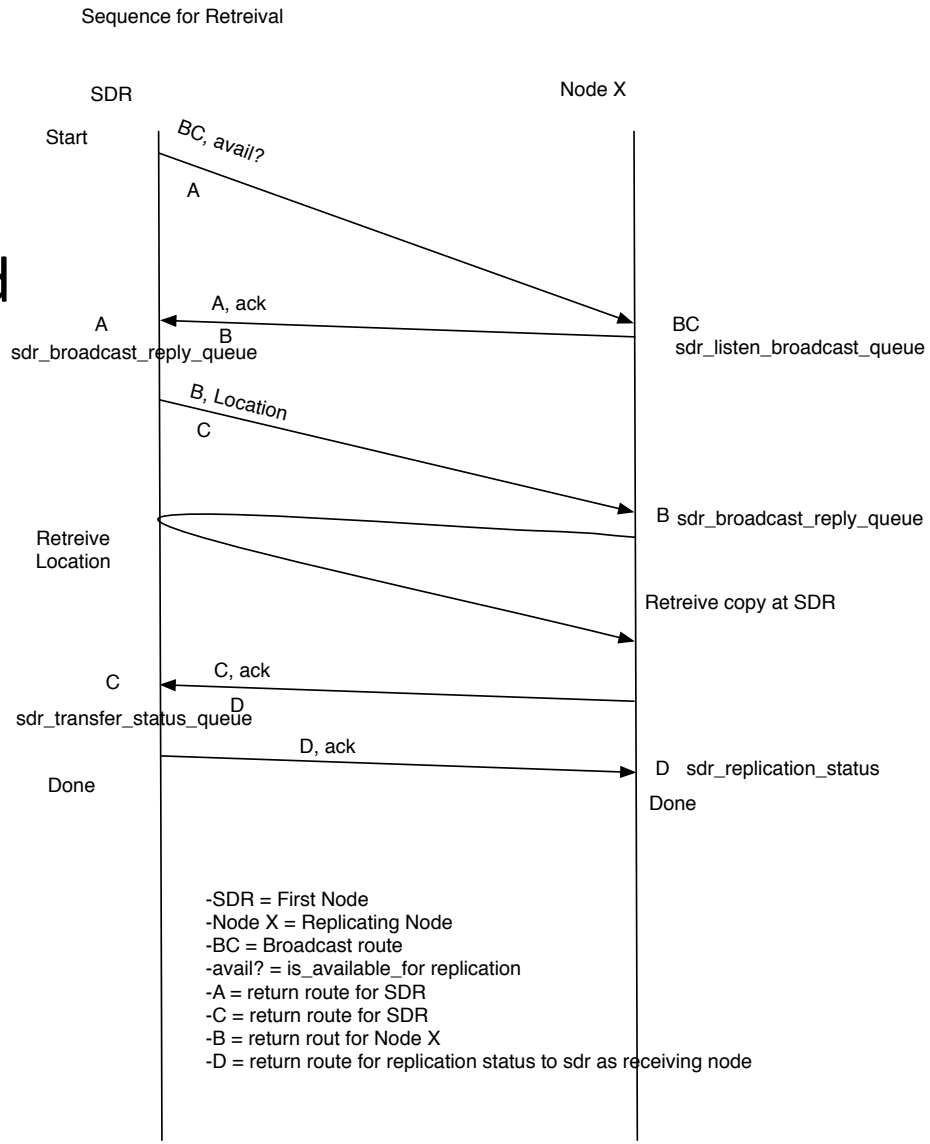
- Replication Request
- Registry Item Create
- Registry Synchronize
- Recovery (digital object, registry entry, registry, etc.)
- Fixity Audit flows
- At any given time each node may be handling multiple message control flows/ sequences at once

Messaging / Not Workflow

- Because DPN is federated and each node is independent,
- DPN does not have a “workflow” system
- So, each of the message sequences provides the mechanism to control the flow of work that needs to be done across the DPN federation
- In this environment, messages are not enough, as there may be failures at any point in a message sequence
- Each node will have to keep track of state of the messages and will have a mechanism to time out and/or recover

Replication Control Flow

- Retrievals happen out of band over secure channel
- Each step can be canceled and the transaction stopped
- Content is checked for fixity (sha256) at the replication node and also at the First node (SDR in the example) after copy
- State of transaction managed by each node separately

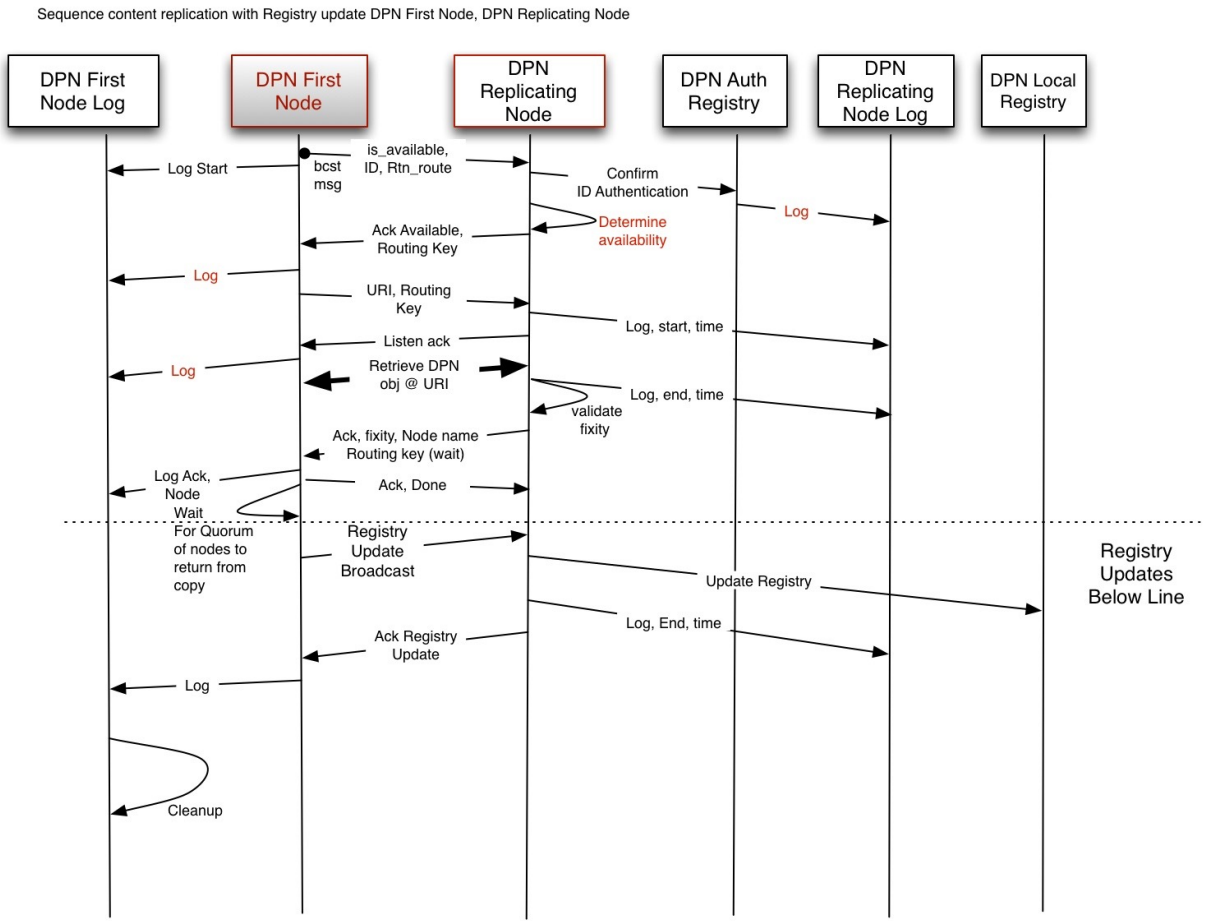


Replication & Registry

- The previous slide showed the first part of a Replication, once a node has copied the content, it must now wait for a message indicating that it must update its registry
- The First Node must also track the successful replications and when enough have completed, issue a Registry Create Entry message to ALL the nodes
- Once the Registry is updated DPN has a copy of the content
- As part of the DPN processes, the DPN messaging protocol must handle partial replication scenarios (e.g. two out of three complete) and incomplete registry updates

Replication With Registry Update and Logging

- A more complete replication entails a few more house keeping steps
- Along with a DPN wide registry update



Registry (messages)

- Currently we are investigating messages that will support Registry services
  - Create, read, update, delete
- Delete is a special case, with special handling
- Creation of new Registry entry will be at the request of a First Node.
  - It will only happen after a quorum of correct copies have been made to Replicating Nodes
  - The Registry entry will be updated at ALL nodes
  - Note that this is a distributed environment, so we expect that the registries will be eventually consistent following Brewers theorem

Registry Item Creation Message


```
{
  "src_node"    : "sdr",
  "message_type": {"broadcast" : "registry_directive"},
  "message"     : "registry_item_create",
  "message_args": [{"dpn_uuid"         : "f81d4fae-7dec-11d0-a765-00a0c91e6bf6",
                    "alt_id"           : "sdr:1234567890"
                    "sha256"           : "2cf24dba5fb0a30e26e83b2ac5b9e29e1b161e5c1fa7425e73043362938b9824",
                    "first_node"       : "sdr",
                    "replicating_nodes": ["ht", "chron"],
                    "previous_version" : "a21d4fae-7dec-11d0-a765-00a0c91e6bcc",
                    "rights_md_ref"    : "c01d4fae-7dec-11d0-a765-00a0c91e6be3"
                  }],
  "date"        : "2013-01-18 09:49:28 -0800"
}
```


- Body of the message needs enough information so that all of the Nodes can track a DPN objects origin and replicating nodes
- Not all nodes that have this entry in their Registry will have copies of the content

DPN Packaging (Bagit)

- We have selected BagIt
- Standard packaging method shared by all nodes
- Minimal, standard bag metadata to enable tracking identity, source and fixity of content bags
- No DPN-wide requirements on descriptive metadata or content structure below the top level bag

DPN Versioning

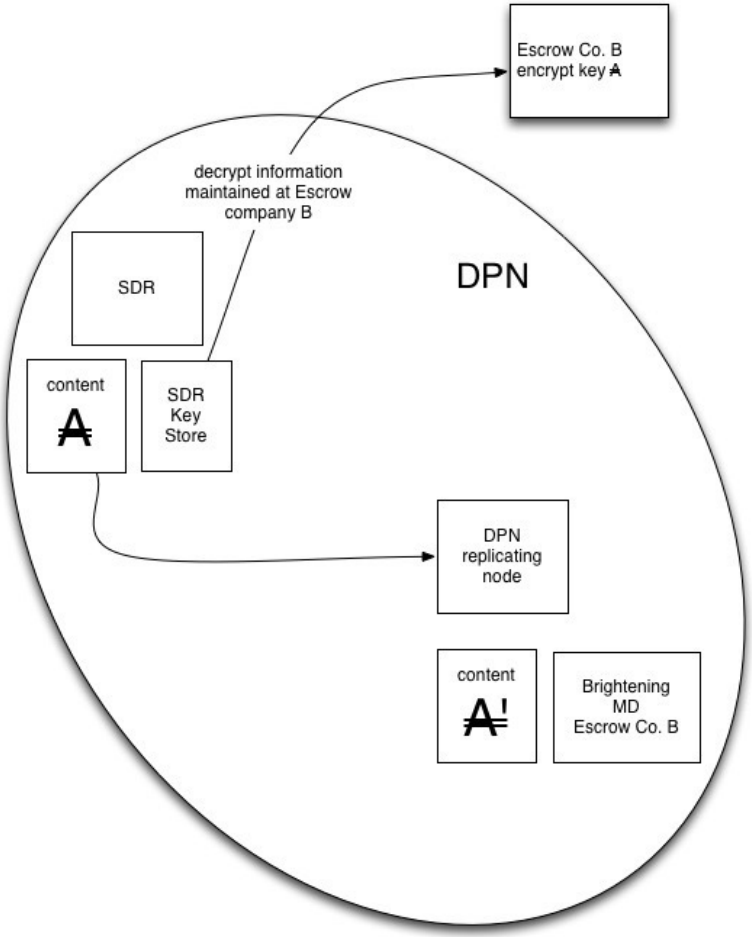
work in progress

- Many of the objects held by DPN will benefit from versioning.
  - First, the archival objects will have a way of creating differential versions, as well as a whole copy version (this may be repository dependent)
  - Second, a number of DPN administrative objects will also benefit from versioning. These objects include Brightening information, Rights Information, Agreements

DPN Encryption

work in progress

- Some content may be encrypted at rest
- Depositors / First Nodes must have confidence that content is secure
- Key escrow to allow content to survive any succession events



DPN Technical Workshop @
PASIG, May 21 in Washington DC

Collaboration Half-Day, 1 - 5 PM
A great chance to have in-depth discussions!