

Academic Preservation Trust

Open Repositories 2013

Scott Turnbull @streamweaver - APTrust

Robert Cartolano - Columbia University

Mission

The Academic Preservation Trust (APTrust) consortium is committed to the creation and management of a preservation repository that will aggregate academic and research content from many institutions.

APTrust Partners

Duke University

Emory University

Johns Hopkins University

University of Maryland

University of North Carolina

N. C. State University

University of Virginia

Columbia University

University of Michigan

University of Notre Dame

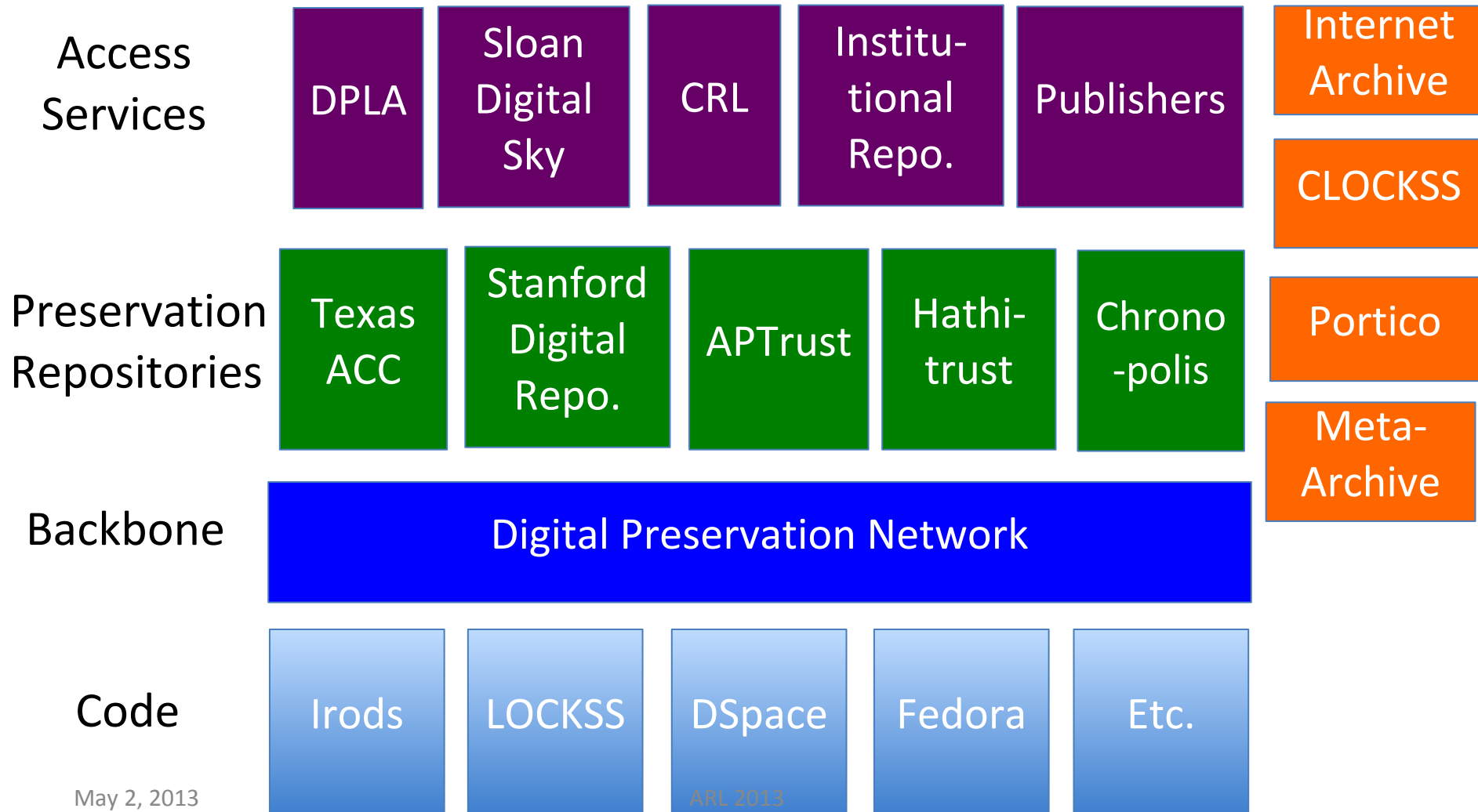
Stanford University

Syracuse University

Development Partner:

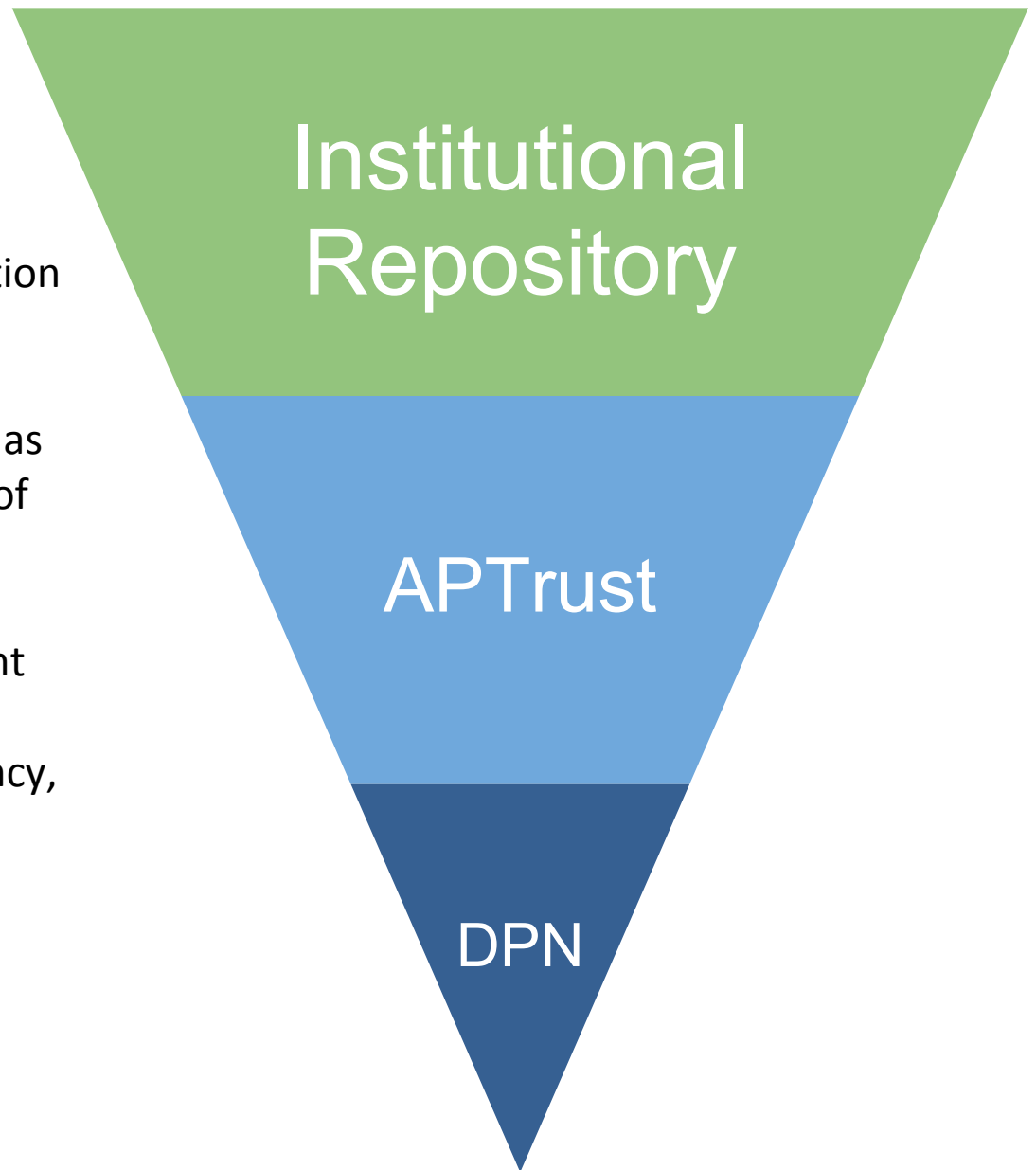
Duraspace

The Emerging Digital Preservation Stack



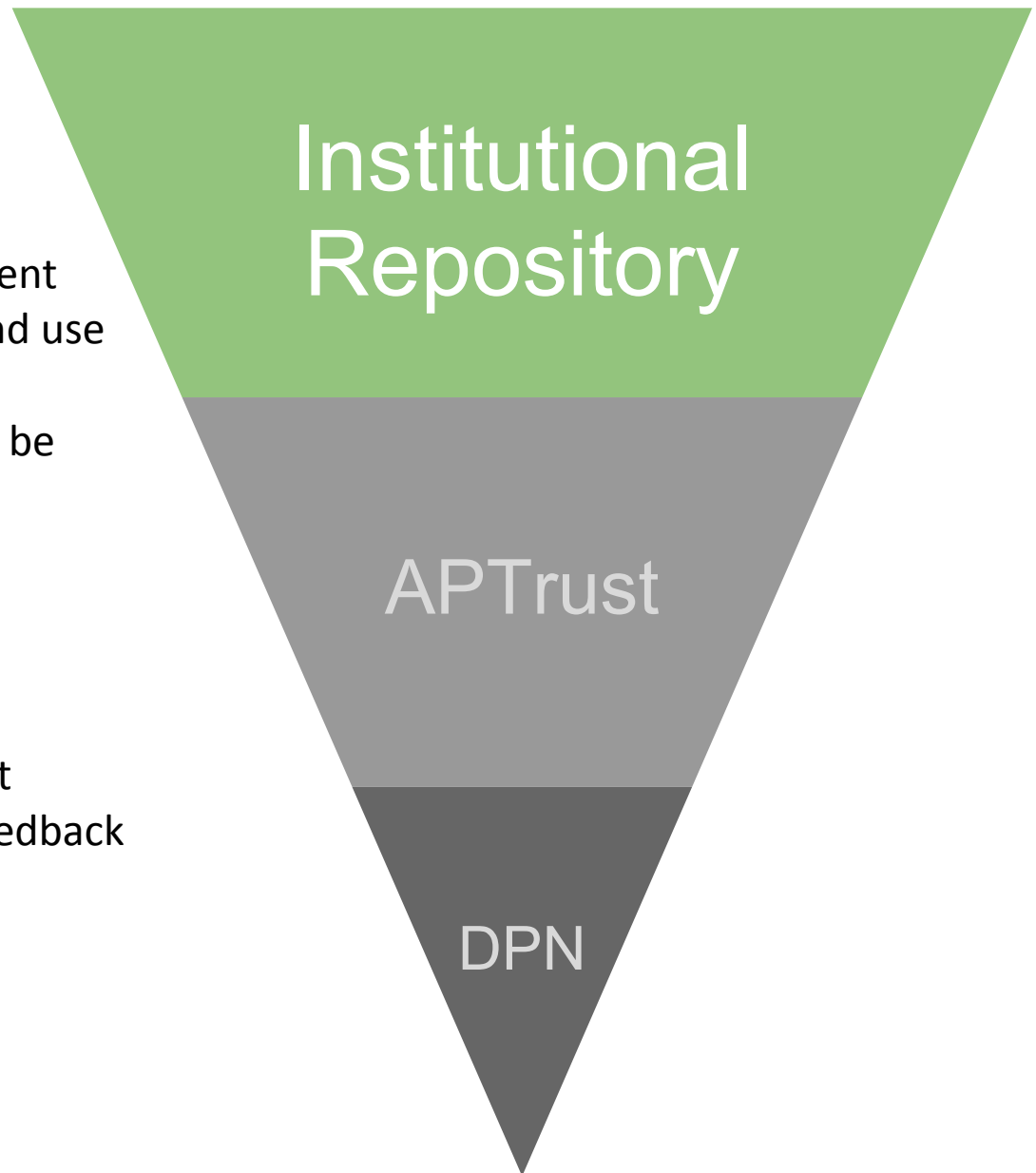
A Continuum of Preservation Services

- Increasing levels of preservation services along NDSA preservation levels
- Winnowing down of content as it passes through each layer of preservation
- Connected services and reporting to help with content management
- Increasing levels of redundancy, geographic diversity and durability



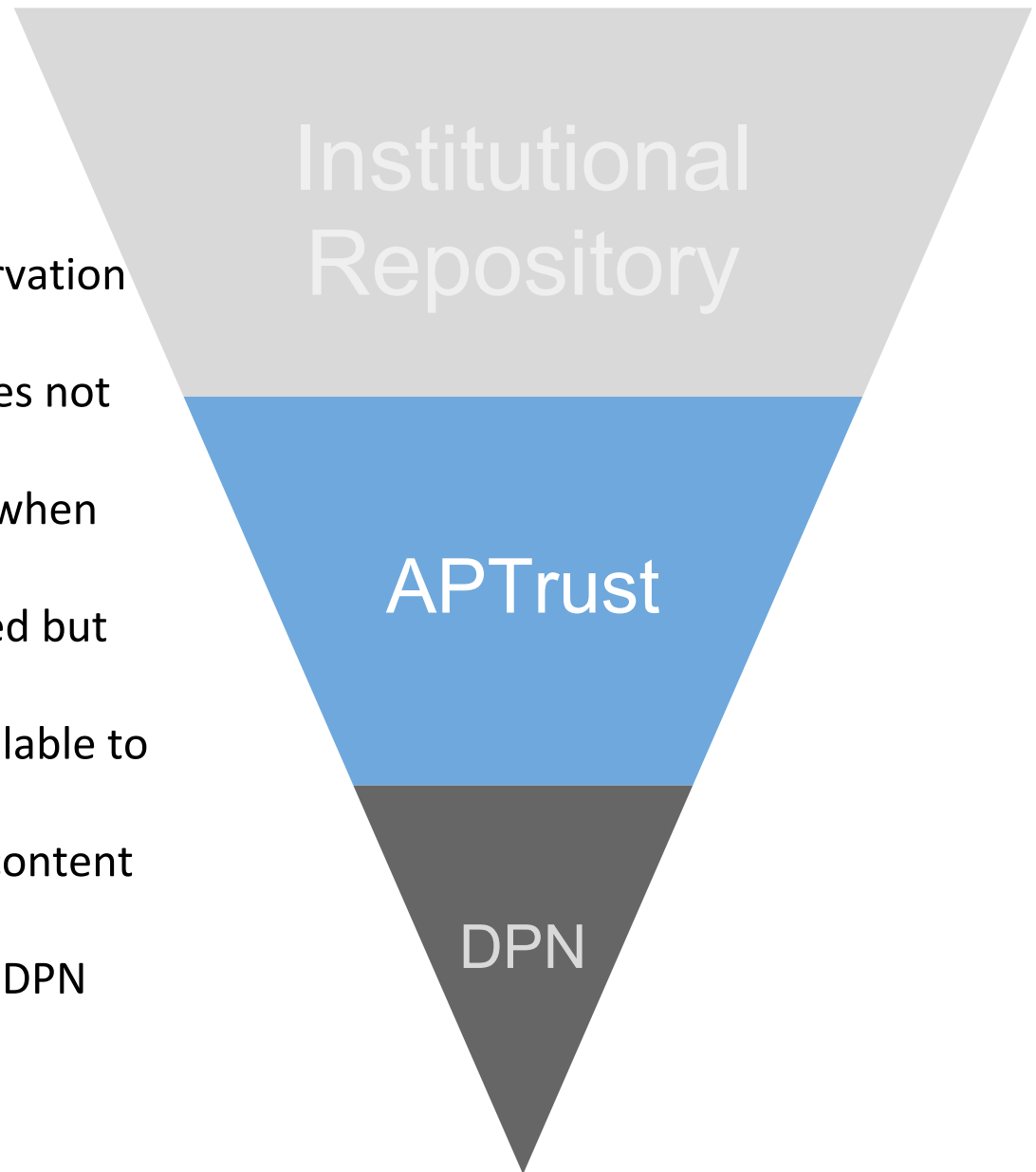
Institutional Repositories

- Producing and Curating Content
- Primary point of discovery and use for their end users
- Full body of content may not be sent to APTrust:
 - Use copies
 - Redundant derivatives
 - Composite works
- Maintain full control and management of their content
- Workflows from sublevels feedback via APIs for reporting and management



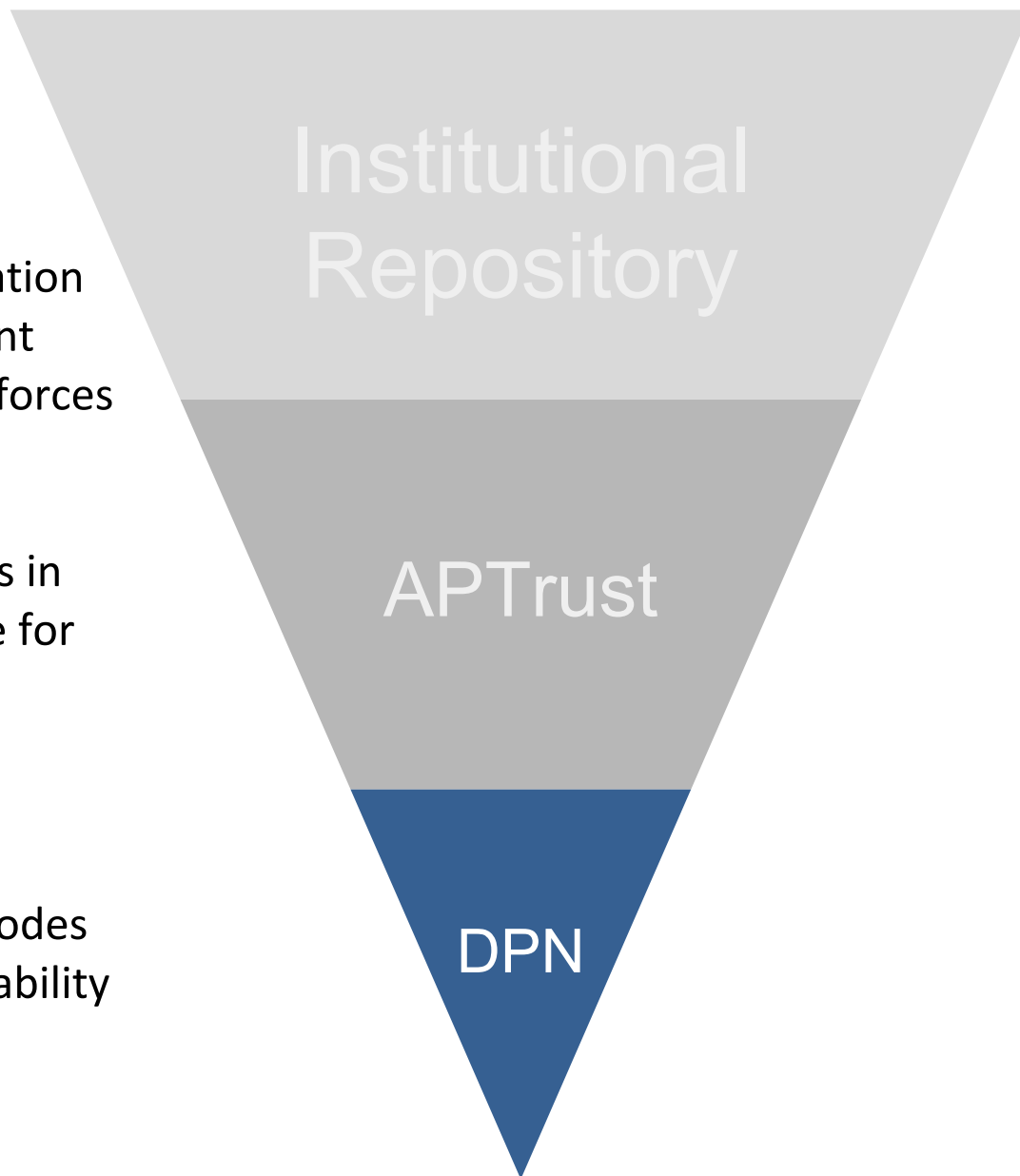
APTrust

- Focuses primarily on preservation
- Proper chain of custody
- Preserving what is sent, does not force a versioning policy
- Receives updates from IRs when they decide
- Allows content to be deleted but will leave a tombstone
- Reporting and Services available to IR via APIs
- Any supplemental data or content sent to institution
- Mediates interactions with DPN

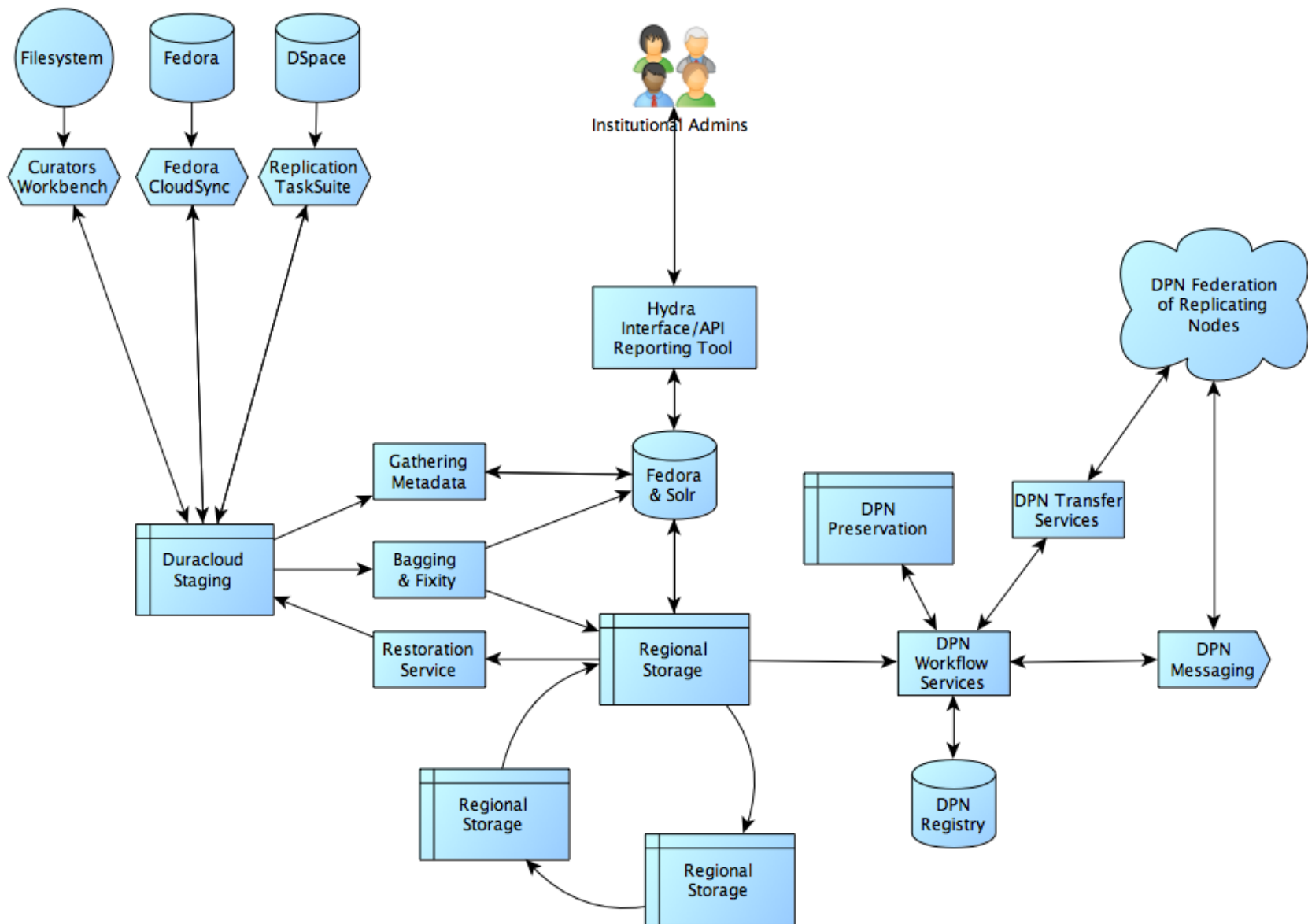


DPN

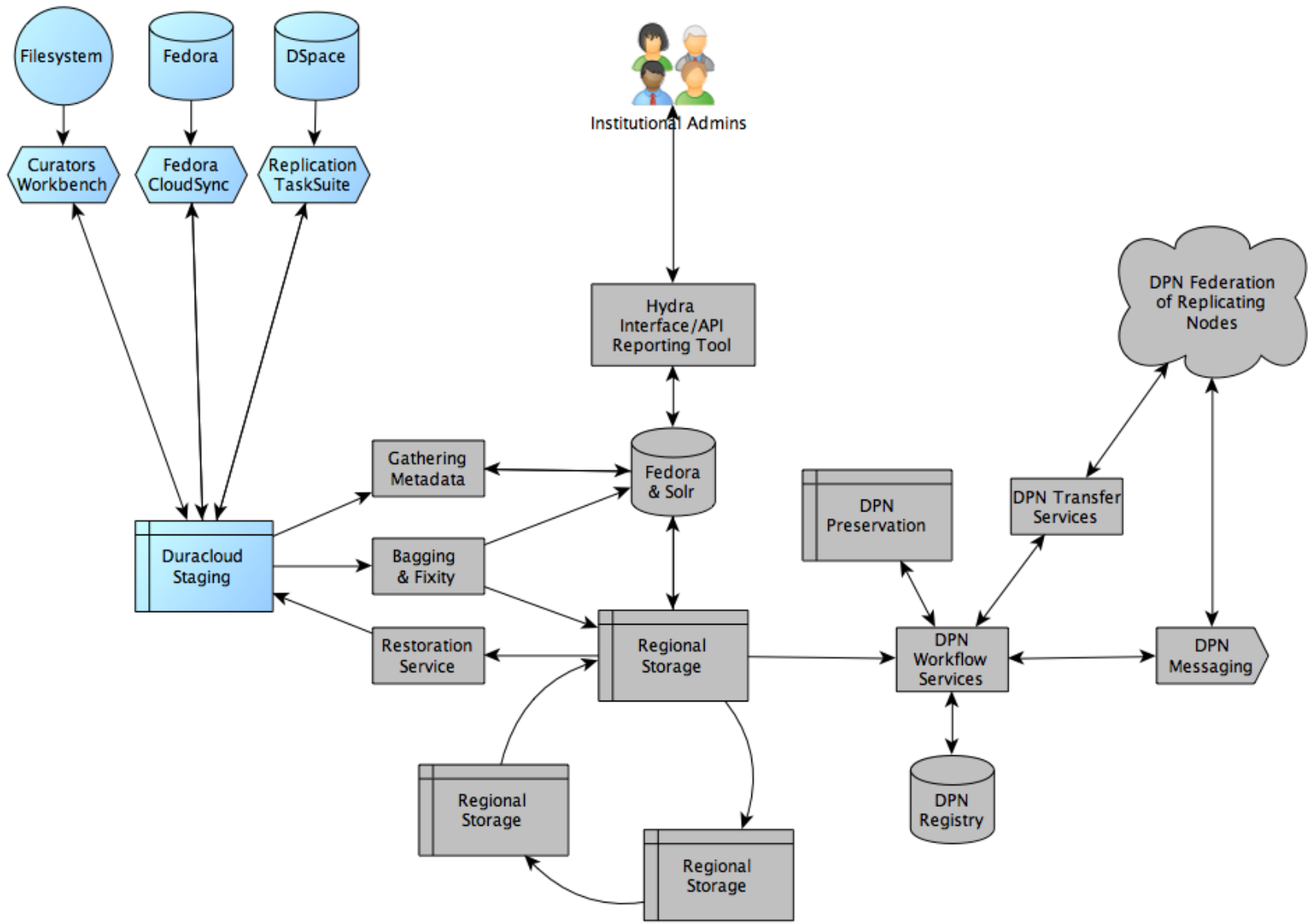
- Focused on critical preservation needs for very stable content
- Can update content and enforces versioning
- No content deletion
- Provides succession services in cases of catastrophic failure for either IR or APTrust
- Secure Dark Archive
- Reporting and interaction mediated through APTrust
- Federation of Replicating Nodes to provide high level of durability



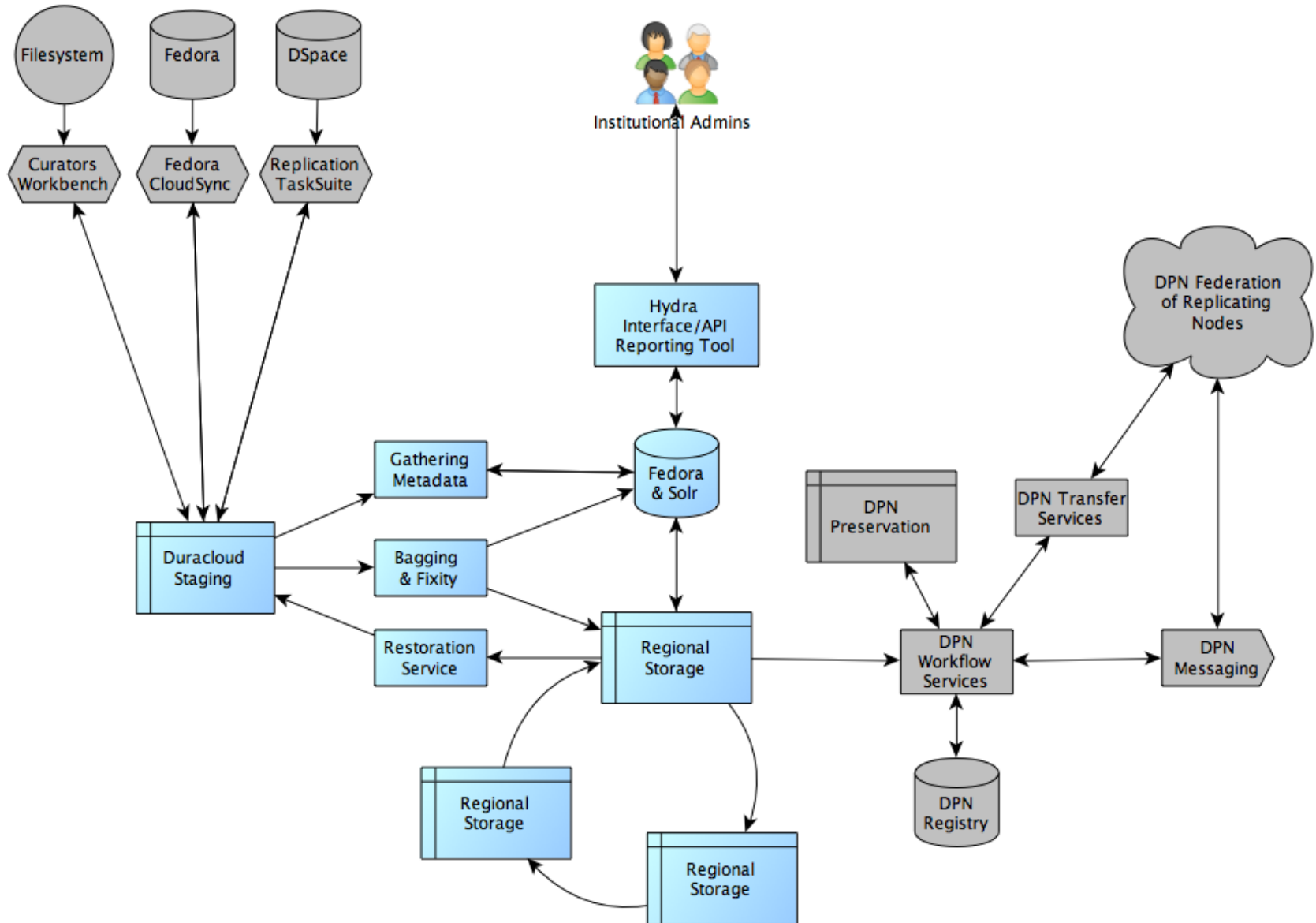
Overall Architecture



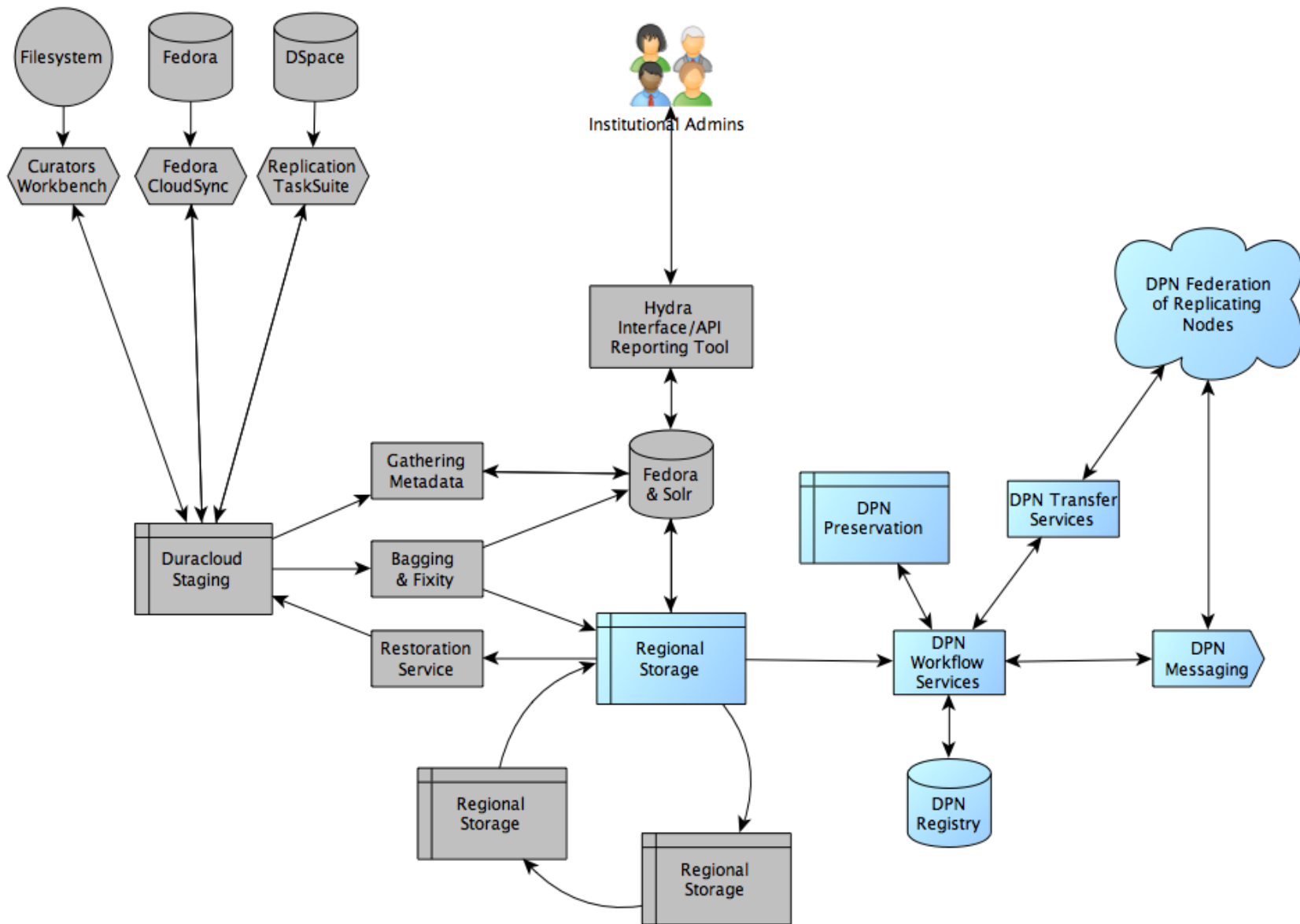
Staging Content for Ingest



Ingest and Manage Content



Sends and Recieves DPN Content



View of Object in Staging to be Bagged

DSpace AIPs (ReplicationTaskSuite)

aip_store_ITEM@123456789-1003.zip

- bitstream_12345.pdf
- bitstream_12346
- mets.xml

Fedora Datastreams (Fedora Cloudsync)

uva-lib:2070291

uva-lib:2070291+RELS-EXT+RELS-EXT.0

uva-lib:2070291+content+content.0

uva-lib:2070291+descMetadata+descMetadata.0

uva-lib:2070291+solrArchive+solrArchive.0

uva-lib:2070291+solrArchive+solrArchive.1

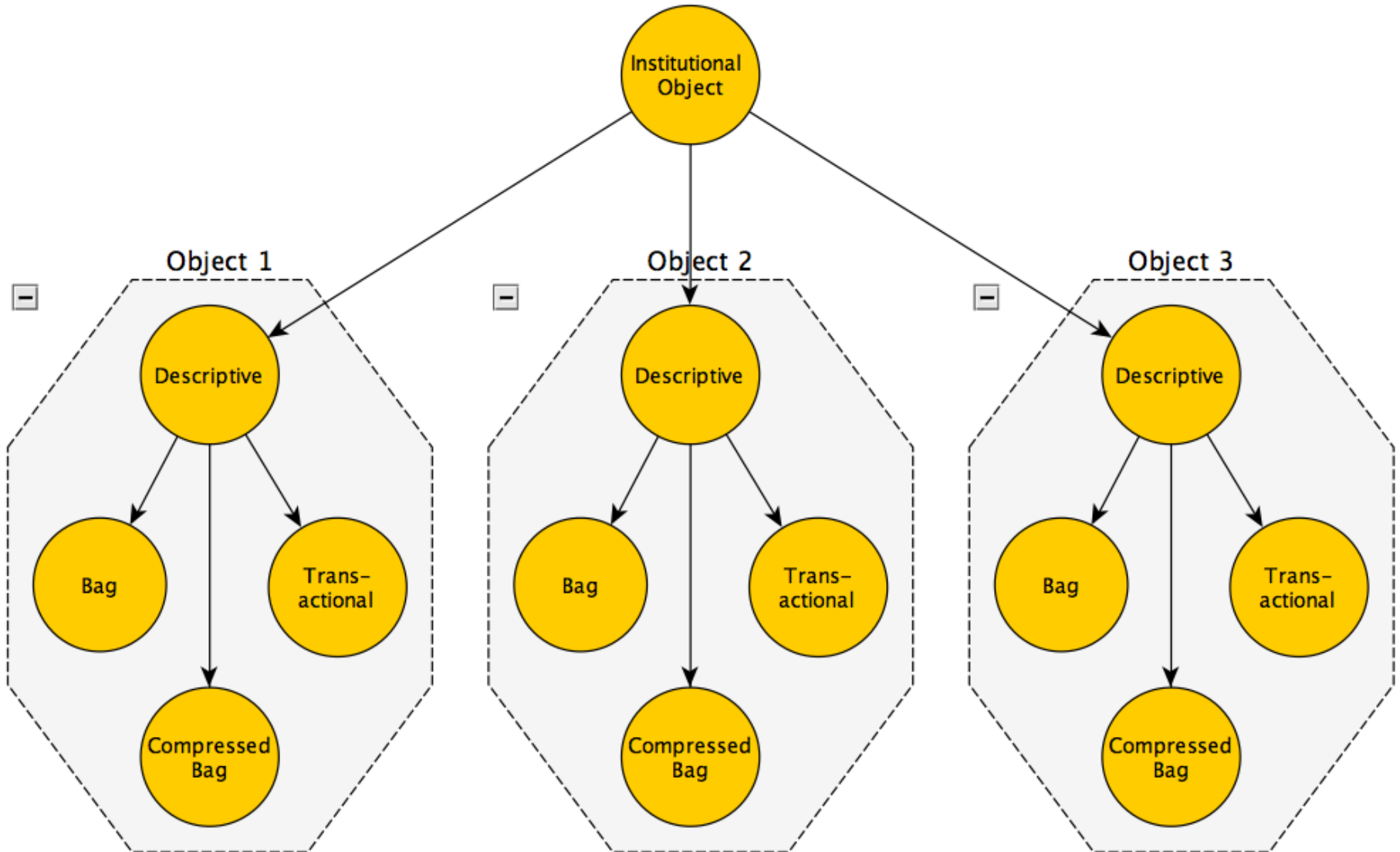
uva-lib:

2070291+technicalMetadata+technicalMetadata
.0

Fedora 4: The Future is Now

- Aiming to launch under Fedora 4
- Configurable storage of great advantage for our use case
- Object Hierarchy (really graph) well suited for managing multi-institutional content
- Clustering and Scalability significantly improved
- Sequences allow processing of content over time and avoiding some ingest bottlenecks

Hierarchical Object Structure

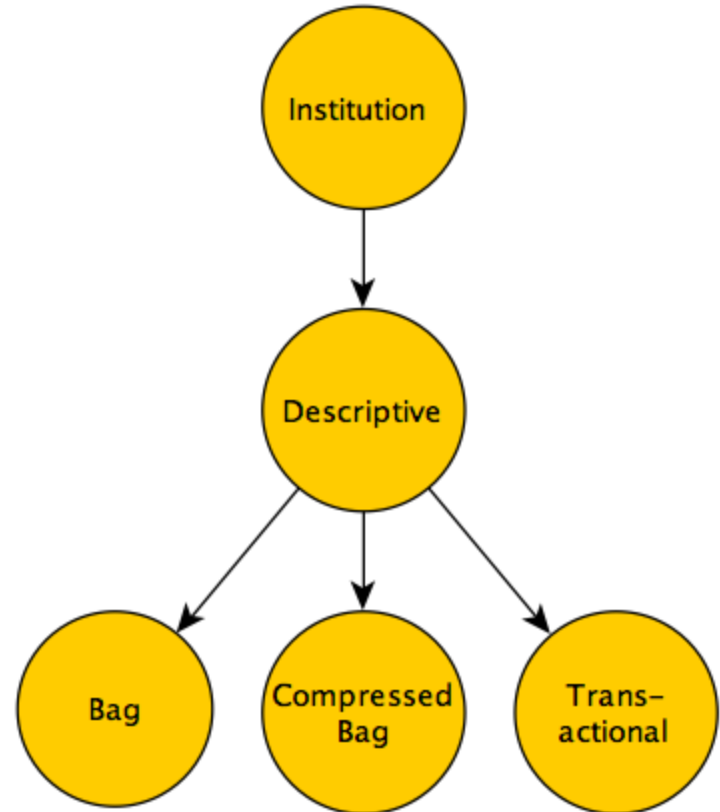


On Twitter: @aptrust

Tweet using: #aptrust

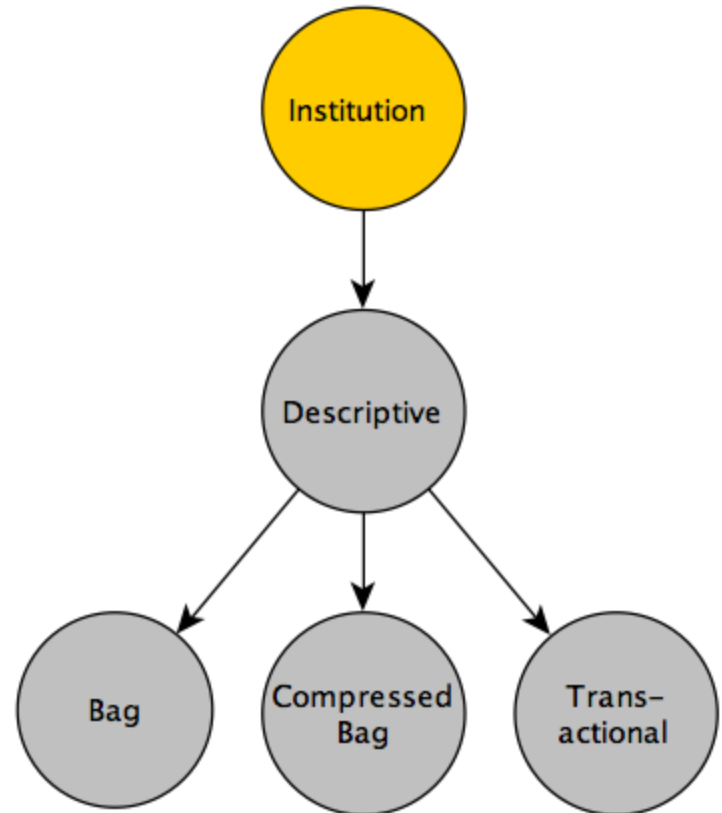
Objects as a Collection of Nodes

- Each object actually a hierarchy of nodes
- Each node serves a specific preservation purpose
- Node structure allows for high level of flexibility in constructing an object



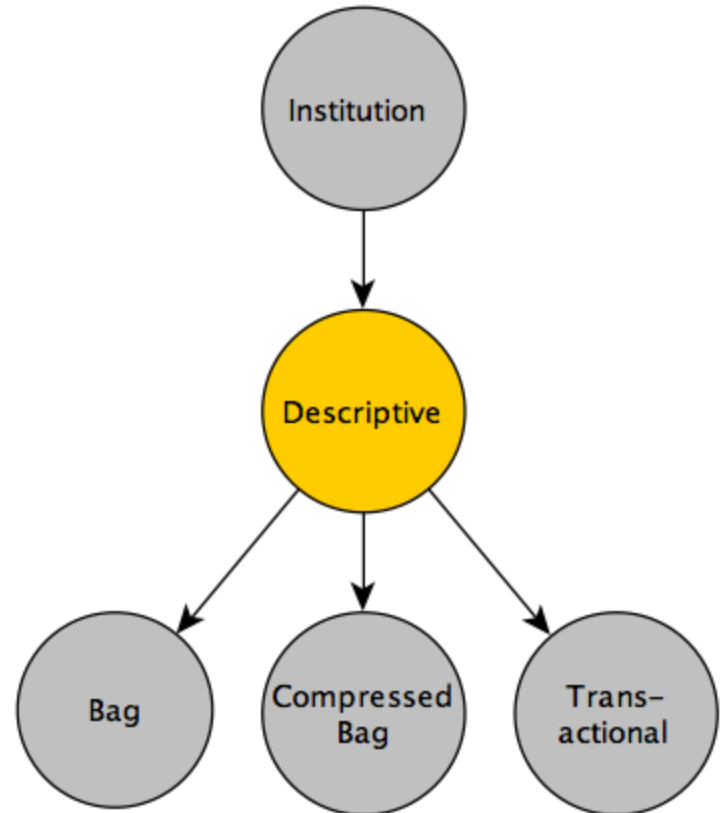
Institution Node

- Maintain metadata about owning institution
- Inform access control to digital objects they own
- Hierarchical Object PIDs mean the Institution is part of object identity
- Disambiguation and collision avoidance



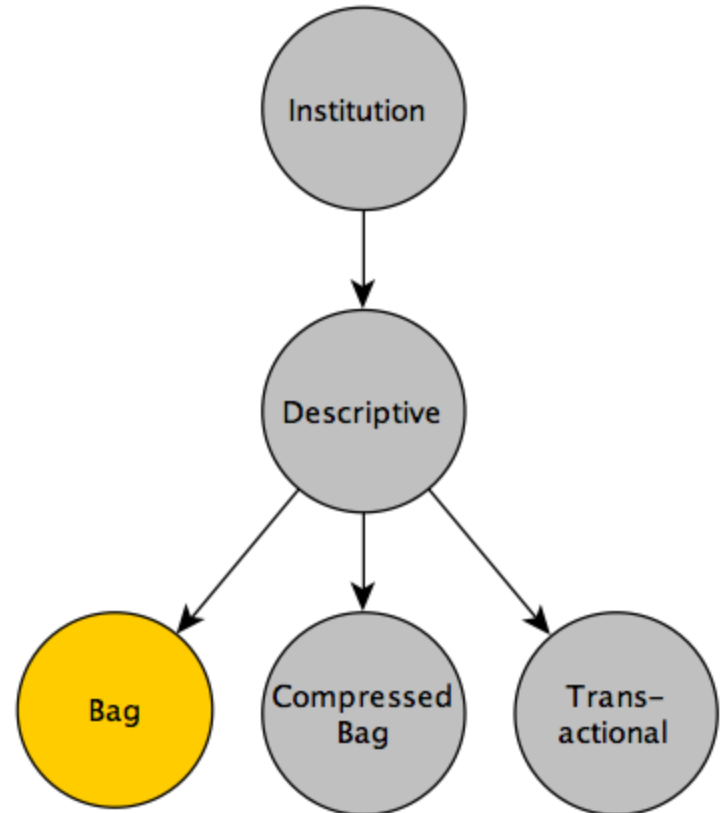
Descriptive Metadata

- Metadata about the object and how to manage it
- Derived from bags on ingest, added via API or both
- Manages Provenance Metadata
- Maintains versioned Metadata
- Persists, even if underlying object deleted



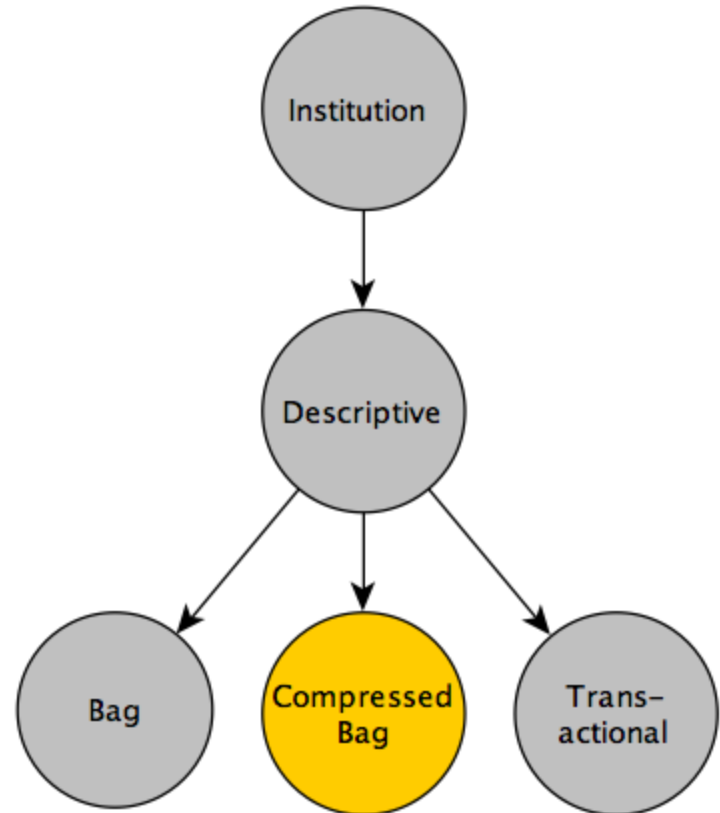
Bag Object

- Generated by processing items from Staging
- Focused on chain of custody and initial preservation
- Initiates sequence to generate other storage nodes
- Used in restoration services to return what was sent
- Can shift to low io storage
- Provides additional durability for content



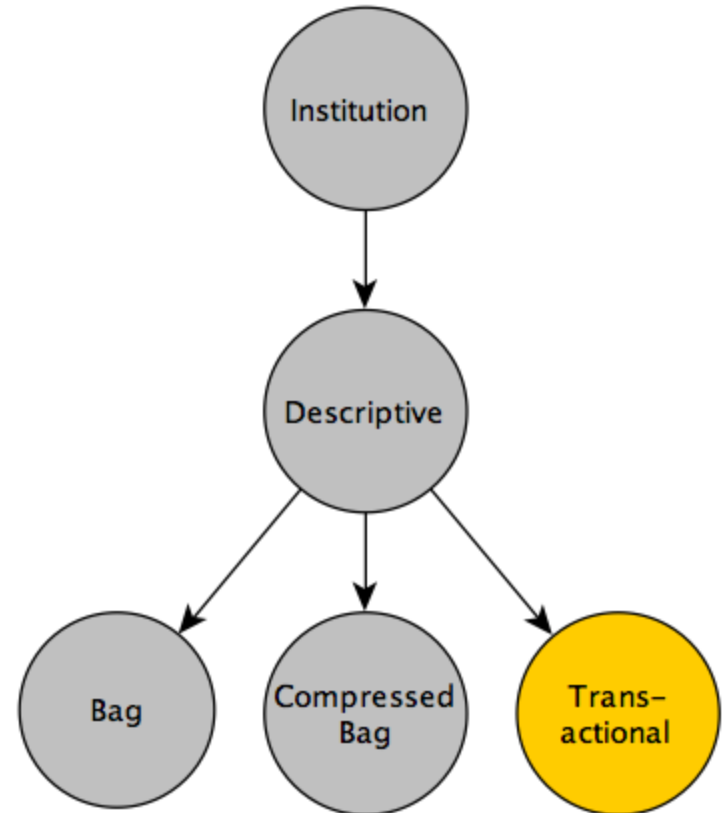
Compressed Bag Object

- Copy of last resort
- Focused on long term and low i/o storage
- Validating compression before considering object final



Transactional Item

- High availability and i/o
- Used for indexing and building services
- Restitched versions of objects if they were chunked
- Used to generate possible use copies or format migration



Collaborative Model

- Owned by the Academy means a focus on collaboratively forming:
 - Governance Model
 - Financial Model
 - Prioritizing development of services
- Leveraging common skill-sets and tools:
 - Positioning partners to collaborate
 - Building opportunities to collaborate

Building Communities & Practice

- UVa/APTrust hosted HydraCamp in early August.
- Bagins - BagIt Library initial release
 - JSON-RPC server goal for this month.
- Provide examples and use cases for Fedora 4 to help build familiarity
- Desire to move quickly to services for enhanced workflows and management

The Year Ahead

- July – Sept:
 - Create Bags from landing space.
 - Establish basic management interface and API.
- Sept – Nov:
 - Object storage configurations and sequences
 - Creation of Transactional & Compressed Objects
- Nov – Dec:
 - Performance improvements
 - Testings and Bug Fixes.
- Jan -> Early 2014
 - Identify and prioritize additional services with partners
 - Begin sending content to DPN

Website: <http://aptrust.org/>

Twitter: <https://twitter.com/APTrust>

GitHub: <https://github.com/APTrust>

scott.turnbull@aptrust.com

Questions?

On Twitter: [@aptrust](https://twitter.com/aptrust)

Tweet using: [#aptrust](https://twitter.com/aptrust)