APTrust Planning

June 19, 2012
APTrust Givens

1. Preservation repository for aggregated content; use of Fedora, Hydra, DuraCloud OS and CloudSync OS technologies

2. DPN replicating node; use of applicable OS technologies

3. Access services:
   1. Phase 1 – administrative access only
   2. Phase 1+ - richer access services determined by APTrust partner community
Aggregate Preservation Repository

1. Accepts all content types
2. Accepts Fedora and non-Fedora content
3. Supports basic ingest & retrieval from multiple institutions
4. Backed up (geographic redundancy)
5. Includes audit mechanisms (goal of trusted repository status in future)
6. “Dusk” archive – administrative access by each institution, potential for end-user access in future
7. Ability to delete content not bound for DPN
DPN replicating node

1. Uses Fedora, DuraCloud OS and CloudSync OS technologies (architectural differentiation from other nodes)
2. Interaction with other DPN nodes (transport mechanisms developed by DPN project)
3. "Dark" archive
4. Roach motel: no deletion
Access Services

Phase I

- “stuff in, stuff out” (mime-typed content)
- Ingest & retrieval, no end-user access
- Audit, periodic and on demand
- Permit future development of more advanced access services
Architecture Guidelines

• Leverage production services
• Minimize modifications and/or new development in the first 18 months
• Align implementation with APTRust Partner Priorities
• Minimize project roadmap dependencies wherever possible (e.g. Hydra, Fedora, potential Cloud providers)
Goals for development sessions

- Review architecture plan
- Test assumptions about technologies and make necessary modifications to plan
- Identify development work and update project plan
- Set milestones for deliverables
- Develop statement of work and timelines
- Issue contract for DuraSpace work
Ingest/Retrieval Workflow & Infrastructure

Legacy Workflow

Preserve External?

Create APTrust Metadata

Stage for Normalization

APTrust Normalization Utility (create/strip APTrust FOXML)

APTrust Storage w/ DuraCloud OS (shib suth, acct, audit, ...)

CloudSync

Staging for Transport & Retrieval

At home institution

Local

APTrust
APTrust DPN Node w/DuraCloud OS (acct, audit, …)

APTrust Storage w/ DuraCloud OS (shib auth, acct, audit, …)

Fedora Aggregation - FOXML & Metadata, externally managed DS

CloudSync

CloudSync *

* work needed to ingest into Fedora and backup/sync into cloud
Interfaces

- Simple Search & Discovery*
- Fedora - FOXML & Metadata, externally managed DS
- Aggregation Storage w/ DuraCloud OS (acct, audit, ...)
- Fedora API – discovery & edits
- DuraCloud API – audit & retrieval
- Admin

* Hydra
### Rough Timeline

<table>
<thead>
<tr>
<th>Date</th>
<th>Milestone</th>
</tr>
</thead>
<tbody>
<tr>
<td>July</td>
<td>Test content prepared</td>
</tr>
<tr>
<td>July</td>
<td>Institutional ingest workflow developed</td>
</tr>
<tr>
<td>July</td>
<td>DuraCloud OS set up (on purchased cloud)</td>
</tr>
<tr>
<td>August</td>
<td>Fedora repository set up (on purchased cloud)</td>
</tr>
<tr>
<td>August</td>
<td>CloudSync installed and tested</td>
</tr>
<tr>
<td>September</td>
<td>Content run through system &amp; tested</td>
</tr>
<tr>
<td>Sept 30</td>
<td>Demo w/out admin interface ready</td>
</tr>
<tr>
<td>Oct-Nov</td>
<td>Develop administrative interface</td>
</tr>
<tr>
<td>Dec 31</td>
<td>Demo w/ admin interface ready</td>
</tr>
</tbody>
</table>
Stuff to be worked out

- Local cloud infrastructure (will begin on Amazon RackSpace)
- Policies governing services (e.g. versioning)
- Business model
- Size limits – especially for data & video
- Replicating node implementation