

APT TRUST TECHNICAL UPDATE

May 5th 2014 Partner Meeting
Columbus, Ohio

Outline of Update

- Progress On Goals
 - Purchasing & Provisioning Infrastructure
 - Receiving & Processing Content
 - Admin & Reporting Interface
 - Preserving Content & Ongoing Services
 - Content Restoration
 - DPN Node
- Next Few Months

Goal Progress: Infrastructure

- Early Feb – DuraSpace Contract for AWS
- Mar 10 – Access to AWS account
- Mar 31 – S3 Buckets Provisioned to Accept Bags
- Currently – Setting up Hydra/Fedora for initial install and partner testing of Admin App
- Status:
 - Able to receive bag content from beta testers
 - Expect initial install and partner access in next few weeks
 - Move to scripted deployments and management during summer

Findings: Infrastructure

- Slower than expected
 - Procurement
 - Setup and installs
- Variable costs per month different than what we're use to
- Ability to regularly pay critical to service stability
- Regular significant price reductions
- Significant misconceptions about cloud services
- Cost estimates holding after initial month of charges

Goal Progress: Processing Content

- UVA, NCSU, Columbia, Miami, Chicago all tested sending bags.
 - Cincinnati & MD preparing to send.
- 233 GB/~5000 bags tested so far
- Orienting partners to tools and packaging format
- Currently able to process valid APTrust Bags
 - Need to mature error handling
- Pending: Register metadata through API
- Pending: Moving content to preservation storage
- Status:
 - Overall in good shape
 - Pending some improvement of architecture

Findings: Processing Content

- Network performance very good
- Very good horizontal scaling
- Anxiety over tools and a place to start
 - Very good feedback on ease of sending once done
- Go language very promising for developing in the cloud
- Model of sync tools vulnerable to version dependencies
- Direct dumps of data can include inappropriate metadata
 - Critical need for self contained “knowledge objects”

Goal Progress: Admin Interface

- Initial admin interface developed with DCE Jan/Feb
 - Roles and access
 - Improved metadata models
 - Basic API (but needs refinement)
- Defined and documented events model for auditing
- Status
 - Need Help! Improve API Authorization model
 - Stuck! Refactoring Routes (Hydra/Rails) – Needed for Update/Delete and other API calls

Findings: Admin Interface

- Performance questionable on Fedora, need more testing with concurrent requests
- Having Hydra community has provided valuable – has limits
- Authorization models are complex and need refinement
- Lots of extra data stored in Fedora

Goal Progress: Preservation Storage & Ongoing Content Management

- Blocked by API and Routes issues
- Next critical step
- Moving processed content to preservation
- Increased resilience by registering some metadata with storage
- Status:
 - Expected to address ~ May/June
 - Still need to detail reasonable way to request deletions

Goal Progress: Restoration

- S3 buckets to send content back established
- Blocked until Reporting and Preservation completed
- Will return an APTrust formatted bag with same content
- APTrust formatted bag also the DPN restoration format
- Status:
 - Pending development expected June/July

Goal Progress: Build DPN Node

- Working with contractors to catch-up
- Expect to test exchanging content with nodes in a few weeks
- Will miss very early tests in June, expect to participate in August tests
- Status
 - Finishing initial ingest process in next few weeks
 - Move to implementing registry creation and sync ~ June
 - Move to implement content restoration ~ July
 - Initial node tests end of August

Findings: DPN Node

- Very loose goals make it difficult to implement
- Significant gaps
 - Deposit agreements
 - Understanding/Defining brightening
 - Undefined service level agreements that may impact our design
 - Content encryption vs no-delete policy
 - Some very basic preservation policies still unresolved
- Lightest staffing of all the nodes

Summary of Next Steps

- Complete installation and setup of Fedora server
- Increase amount of test content
- Refinement of API
- Clear routes block
- Register processed metadata through API
- Store content in preservation space and implement scheduled preservation activities
- Targeting early August for initial preservation
- Timeline dependent on partner involvement