

# Academic Preservation Trust

**Author:** The DSpace Developer Team  
**Date:** 09 February 2017  
**URL:** <https://wiki.duraspace.org/display/aptrust>

# Table of Contents

1	This wiki has moved! Please see <a href="https://sites.google.com/a/aptrust.org/aptrust-wiki/">https://sites.google.com/a/aptrust.org/aptrust-wiki/</a>	12
1.1	Search the APTTrust wiki	12
1.2	General	12
1.3	APTrust Implementation April 2013	12
1.4	Development Best Practices	13
1.4.1		13
1.4.2	Meetings & Calls	13
1.4.3	Page tree	13
2	APTrust Conference Line	15
2.1	Phone Instructions	15
2.2	Skype Instructions	15
3	Institution Codes	16
4	APTrust Meeting Notes	17
4.1	2012-08-06 - APTrust Developer Call	17
4.1.1	Attendees	17
4.1.2	Discussion	17
4.2	2012-08-20 - APTrust Developer call (by Donna Tolson)	19
4.2.1	Attendees	19
4.2.2	Absent	19
4.2.3	Discussion	19
4.3	2012-08-20 - APT Strategy Meeting (by Andrew Woods)	21
4.3.1	Attendees	21
4.3.2	Call-in Details	21
4.3.3	Discussion	21
4.4	Attendees	23
4.4.1	Absent	23
4.4.2	Discussion	23
4.5	2012-09-24 & 25 - APTrust Face-to-Face Meeting - Charlottesville	24
4.5.1	<code>/*&lt;![CDATA[*]*/ div.rbtoc1486669572852 {padding: 0px;} div.rbtoc1486669572852 ul {list-style: disc;margin-left: 0px;} div.rbtoc1486669572852 li {margin-left: 0px;padding-left: 0px;} /*]]&gt;*/</code>	24
4.5.2	Attendees	25
4.5.3	Location and Time	25
4.5.4	Agenda	26
4.5.5	Architecture	27
4.5.6	Whiteboard	27
4.5.7	Discussion	28
4.6	2012-10-01 - APT Strategy Meeting	40
4.6.1	Attendees	40
4.6.2	Call-in Details	40

4.6.3	Agenda	40
4.6.4	Discussion	40
4.6.5	Actions	41
4.7	2012-10-09 - APT Admin Console Meeting	42
4.7.1	Attendees	42
4.7.2	Call-in Details	42
4.7.3	Agenda	42
4.7.4	Discussion	42
4.7.5	Actions	43
4.8	2012-10-15 - APT Strategy Meeting	44
4.8.1	Attendees	44
4.8.2	Call-in Details	44
4.8.3	Agenda	44
4.8.4	Status	44
4.8.5	Discussion	45
4.8.6	Actions	46
4.9	2012-10-16 - APT Admin Console Meeting	47
4.9.1	Attendees	47
4.9.2	Call-in Details	47
4.9.3	Agenda	47
4.9.4	Discussion	47
4.9.5	Actions	49
4.10	2012-10-18 - APT Admin Console Meeting	49
4.10.1	Attendees	49
4.10.2	Call-in Details	49
4.10.3	Agenda	49
4.10.4	Discussion	50
4.10.5	Actions	51
4.11	2012-10-30 - APT Admin Console Meeting	52
4.11.1	Attendees	52
4.11.2	Call-in Details	52
4.11.3	Agenda	52
4.11.4	Previous Actions	52
4.11.5	Discussion	52
4.11.6	Actions	53
4.12	2012-11-08 - APT Admin Console Meeting	54
4.12.1	Attendees	54
4.12.2	Call-in Details	54
4.12.3	Agenda	54
4.12.4	Previous Actions	54
4.12.5	Discussion	54
4.12.6	Actions	56
4.13	2012-11-13 - APT Admin Console Meeting	56

4.13.1	Attendees	56
4.13.2	Call-in Details	56
4.13.3	Agenda	56
4.13.4	Previous Actions	56
4.13.5	Discussion	57
4.13.6	Actions	58
4.14	2012-11-13 - AP Trust Technical Update	58
4.14.1	Agenda	58
4.14.2	Format	59
4.15	2012-11-26 - APT Strategy Meeting	59
4.15.1	Attendees	59
4.15.2	Call-in Details	59
4.15.3	Agenda	60
4.15.4	Discussion	60
4.15.5	Actions	61
4.16	2012-11-27 - APT Admin Console Meeting	62
4.16.1	Attendees	62
4.16.2	Call-in Details	62
4.16.3	Agenda	62
4.16.4	Previous Actions	62
4.16.5	Discussion	62
4.16.6	Actions	63
4.17	2012-12-04 - APT Admin Console Meeting	63
4.17.1	Attendees	63
4.17.2	Call-in Details	64
4.17.3	Agenda	64
4.17.4	Previous Actions	64
4.17.5	Discussion	64
4.17.6	Actions	65
4.18	2013-01-08 - APT Admin Console Meeting	65
4.18.1	Attendees	65
4.18.2	Call-in Details	65
4.18.3	Agenda	66
4.18.4	Previous Actions	66
4.18.5	Discussion	66
4.18.6	Actions	67
4.19	2013-01-15 APTrust Admin Console Meeting	68
4.19.1	Attendees	68
4.19.2	Call-in Details	68
4.19.3	Agenda	68
4.19.4	Previous Actions	68
4.19.5	Discussion	69
4.19.6	Actions	70

4.20	2013-01-22 - APTrust Admin Console Meeting	70
4.20.1	Attendees	70
4.20.2	Call-in Details	70
4.20.3	Agenda	71
4.20.4	Previous Actions	71
4.20.5	Discussion	71
4.20.6	Actions	73
4.21	2013-01-29 APTrust Working Group Meeting	73
4.21.1	Attendees	73
4.21.2	Call-in Details	73
4.21.3	Agenda	73
4.21.4	Previous Actions	73
4.21.5	Discussion	74
4.21.6	Actions	74
4.22	2013-02-01 APTrust DSpace Partner Discussion	74
4.22.1	Attendees	74
4.22.2	Call-in Details	74
4.22.3	Agenda	75
4.22.4	Minutes	75
4.23	2013-02-04 APTrust Strategy meeting	76
4.23.1	Attendees	76
4.23.2	Call-in Details	76
4.23.3	Agenda	77
4.24	2013-02-05 APTrust Working Group Meeting	78
4.24.1	Attendees	78
4.24.2	Call-in Details	78
4.24.3	Agenda	78
4.24.4	Previous Actions	79
4.24.5	Discussion	79
4.24.6	Actions	80
4.25	2013-02-14 APTrust Working Group Meeting	80
4.25.1	Attendees	80
4.25.2	Call-in Details	80
4.25.3	Agenda	80
4.25.4	Previous Actions	80
4.25.5	Discussion	81
4.25.6	Actions	82
4.26	2013-02-19 APTrust Working Group Meeting	82
4.26.1	Attendees	82
4.26.2	Call-in Details	82
4.26.3	Previous Actions	83
4.26.4	Discussion	83
4.26.5	Actions	84

4.27	2013-02-25 APTrust Strategy Meeting	84
4.27.1	Attendees	84
4.27.2	Call-in Details	85
4.27.3	Discussion	85
4.28	2013-02-26 APTrust Working Group Meeting	86
4.28.1	Attendees	86
4.28.2	Call-in Details	86
4.28.3	Agenda	86
4.28.4	Previous Actions	87
4.28.5	Discussion	87
4.28.6	Actions	88
4.29	2013-03-05 APTrust Tech Summit Meeting	88
4.29.1	APTrust Tech Summit Notes	88
4.29.2	Monday	88
4.30	2013-03-05 APTrust Working Group Meeting	94
4.30.1	Attendees	94
4.30.2	Call-in Details	94
4.30.3	Agenda	95
4.30.4	Previous Actions	95
4.30.5	Discussion	95
4.30.6	Actions	96
4.31	2013-03-19 APTrust Working Group Meeting	96
4.31.1	Attendees	96
4.31.2	Call-in Details	97
4.31.3	Agenda	97
4.31.4	Previous Actions	97
4.31.5	Discussion	97
4.31.6	Actions	98
4.32	2013-03-21 APTrust Strategy Meeting	99
4.32.1	Attendees	99
4.32.2	Call-in Details	99
4.32.3	Agenda	100
4.32.4	Discussion	100
4.32.5	Actions	102
4.33	2013-03-26 APTrust Working Group Meeting	103
4.33.1	Attendees	103
4.33.2	Call-in Details	103
4.33.3	Agenda	103
4.33.4	Previous Actions	103
4.33.5	Discussion	103
4.33.6	Actions	104
4.34	2013-04-02 APTrust DPN Node Discussion	104
4.34.1	Attendees	104

4.34.2	Call-in Details	104
4.34.3	Agenda	104
4.34.4	Discussion	105
4.35	2013-04-09 APTrust Working Group Meeting	107
4.35.1	Attendees	107
4.35.2	Call-in Details	107
4.35.3	Agenda	107
4.35.4	Previous Actions	107
4.35.5	Discussion	108
4.36	2013-04-16 APTrust Working Group Meeting	109
4.36.1	Attendees	109
4.36.2	Call-in Details	109
4.36.3	Agenda	109
4.36.4	Previous Actions	109
4.36.5	Discussion	109
4.37	2013-05-07 APTrust Working Group Meeting	110
4.37.1	Attendees	110
4.37.2	Call-in Details	111
4.37.3	Agenda	111
4.37.4	Discussion	111
4.38	2013-05-28 APTrust Working Group Meeting	113
4.38.1	Attendees:	113
4.38.2	Discussion	113
4.38.3	Next Steps	114
4.39	2013-06-04 APTrust Working Group Meeting	114
4.39.1	Attendees:	114
4.39.2	Agenda	114
4.39.3	Updates from Sub Groups	115
4.39.4	Discussion	115
4.39.5	Next Steps & Open Issues	116
4.40	2013-06-18 APTrust Working Group Notes	116
4.40.1	Attendees:	116
4.40.2	Agenda	116
4.40.3	Discussion	117
4.40.4	Next Steps & Open Issues	118
4.41	2013-06-25 APTrust Working Group Meeting	118
4.41.1	Attendees:	118
4.41.2	Agenda	118
4.41.3	Discussion	119
4.41.4	Next Steps & Open Issues	119
4.42	2013-07-02 APTrust Working Group Meeting	120
4.42.1	Attendees:	120
4.42.2	Agenda	120

4.42.3	Discussion	120
4.42.4	Next Steps & Open Issues	122
4.43	2013-07-16 APTrust Working Group Meeting	122
4.43.1	Attendees:	122
4.43.2	Agenda	123
4.43.3	Discussion	123
4.43.4	Next Steps & Open Issues	124
4.44	2013-07-23 APTrust Working Group Meeting	124
4.44.1	Attendees:	124
4.44.2	Agenda	124
4.44.3	Discussion	124
4.44.4	Next Steps & Open Issues	125
4.45	2013-07-30 APTrust Working Group Meeting	126
4.45.1	Attendees:	126
4.45.2	Agenda	126
4.45.3	Discussion	126
4.45.4	Next Steps & Open Issues	127
4.46	2013-08-13 APTrust Working Group Meeting	128
4.46.1	Attendees:	128
4.46.2	Agenda	128
4.46.3	Discussion	128
4.46.4	Next Steps & Open Issues	129
4.47	2013-08-20 APTrust Working Group Meeting	129
4.47.1	Attendees:	129
4.47.2	Agenda	129
4.47.3	Discussion	130
4.47.4	Next Steps & Open Issues	131
4.48	2013-08-27 APTrust Working Group Meeting	132
4.48.1	Attendees:	132
4.48.2	Agenda	132
4.48.3	Discussion	133
4.48.4	Next Steps & Open Issues	134
4.49	2013-09-03 APTrust Working Group Meeting	134
4.49.1	Attendees:	134
4.49.2	Agenda	134
4.49.3	Discussion	135
4.49.4	Next Steps & Open Issues	136
4.50	2013-09-10 APTrust Working Group Meeting	137
4.50.1	Attendees:	137
4.50.2	Agenda	137
4.50.3	Discussion	137
4.50.4	Next Steps	138
4.51	2013-09-17 APTrust Working Group Meeting	138



4.51.1	Attendees: _____	138
4.51.2	Agenda _____	138
4.51.3	Discussion _____	138
4.51.4	Next Steps _____	139
4.52	2013-09-23 APTrust Working Group Meeting _____	139
4.52.1	Attendees: _____	139
4.52.2	Agenda _____	140
4.52.3	Discussion _____	140
4.52.4	Next Steps _____	140
5	AP Trust Package Submission Metadata _____	141
5.1	Phase 1 Functional Requirements _____	141
5.2	Outstanding Questions _____	142
5.3	APTrust Package Object _____	142
5.3.1	Package Object Creation _____	142
5.3.2	Relationships _____	143
6	Architecture _____	145
6.1	This page is currently being edited to reflect the current development and workflow strawman scenario for group feedback. _____	145
6.2	Current AP Trust Functional Requirements _____	145
6.3	Overview of Flow of Digital Objects _____	145
6.4	Current System Architecture _____	146
6.5	Overview of Components _____	146
6.5.1	Staging Of Content for Ingest _____	146
6.5.2	Content Processing _____	147
6.5.3	Administrative Interface _____	147
6.5.4	Preservation Space _____	147
6.5.5	Restoration Services _____	147
6.5.6	Sending Content to DPN _____	148
6.5.7	Earlier Architecture Proposals and Documentation _____	148
6.6	AP Trust Administrative Interface _____	148
6.6.1	<code>/*&lt;![CDATA[*] div.rbtoc1486669573549 {padding: 0px;} div.rbtoc1486669573549 ul {list-style: disc;margin-left: 0px;} div.rbtoc1486669573549 li {margin-left: 0px;padding-left: 0px;} /*]]&gt;*/</code> _____	148
6.6.2	Functionality _____	149
6.6.3	Mock-ups _____	151
6.6.4	Backend Needs for Admin Interface _____	154
6.7	APTrust Data Storage Options _____	157
6.7.1	Storage Goals _____	157
6.8	AP Trust Participant Setup _____	158
6.9	Architecture Overview (March 12) _____	159
6.9.1	Thoughts evolving from 21 March meeting at UVa Library (representatives of Duraspace and UVa present): _____	159
6.9.2	APT Architecture Sketches _____	159
6.10	Export Processing Service _____	161

6.11	Ingest Client	161
6.11.1	Ingest Client Data Flow (simple)	161
6.11.2	Ingest Client Data Flow (sophisticated)	162
6.11.3	Functional Requirements	162
6.11.4		162
6.11.5	Ingest Manifest	163
6.12	Ingest Processing Service	165
6.12.1	Functionality	165
6.12.2	REST API	166
6.12.3	Ingest Workflow State	166
6.13	Submission workflow and tools	168
7	Best Practices	170
7.1	Coding Styleguide	170
7.1.1	Overview	170
7.1.2	Ruby	170
7.1.3	Python	171
7.1.4	Java	171
7.2	Pivotal Tracker	172
7.2.1	Overview	172
7.2.2	Using Pivotal Tracker	172
7.2.3	User Stories	173
7.2.4	Story Points & Estimation	174
7.2.5	GitHub Post Commit Hooks	174
7.3	Service Setup and Configuration	175
7.3.1	Setting Up a Server and Secondary Server	175
7.4	Overview	177
7.4.1	Branching and Tagging Strategy	177
8	Project Information	178
8.1	Purpose	178
8.2	APT 1.0 Guiding Documents	178
8.2.1	General Assumptions	178
8.2.2	APT 1.0 Implementation Project Charter	178
8.2.3	APT 1.0 Implementation Project Requirements	178
8.2.4	Communication Plan	178
8.2.5	Navigate space	178
8.3	APT1.0 in Detail	178
8.3.1	APT 1.0 Implementation Project Plan	178
8.3.2	APT 1.0 Implementation Timelines	179
8.3.3	APT 1.0 Roadmap and Outstanding Work	179
8.3.4	APT 1.0 Implementation Project Lessons Learned	179
8.4	APT1.0 Outstanding Work	179
8.4.1	Test Content	181
8.5	APT 1.0 Communication Plan	183

8.5.1	Goals	183
8.5.2	Audience Definition	183
8.5.3	Communication Tools	183
8.5.4	Message, Priority of Communication	184
8.6	Dec, 2012	184
8.6.1	Basic Flow	184
8.6.2	Admin Web Console	186
9	Technical milestones	187
10	Workflow Description Summary	189
10.1	Preservation Package Workflow	189
10.1.1	Packaging	189
10.1.2	Staging	190
10.1.3	Ingestion	190
10.1.4	Export	190

# 1 This wiki has moved! Please see <https://sites.google.com/a/aptrust.org/aptrust-wiki/>

---



The main Academic Preservation website with news and general information is at <http://aptrust.org>. This complementary wiki hosts more detailed information.

## 1.1 Search the APTrust wiki

---

## 1.2 General

---

- [News](#)
- [About APTrust](#)
- [Governance](#)
- [Contact](#)
- [Partners](#)
- Pivotal Tracker
  - [Main APTrust Pivotal Tracker Project](#)
  - [APTrust-DPN Pivotal Tracker Project](#)
  - [APTrust Systems Pivotal Tracker Project](#)
- [APTrust Internal Wiki](#) (You wont need to access this unless you're an APTrust admin)

## 1.3 APTrust Implementation April 2013

---

- [Architecture](#)
- [Workflow Description](#)
- [Project Information](#)
- [AP Trust Package Submission Metadata](#)
- [Phase I milestones](#)
- [APTrust Institutional Codes](#)

## 1.4 Development Best Practices

---

### 1.4.1

- [Coding Styleguide](#)
- [Pivotal Tracker](#)
- [Service Setup and Configuration](#)
  - [Demo Configuration and Setup](#)
- [Version Control](#)

### 1.4.2 Meetings & Calls

- [Meetings](#)

### 1.4.3 Page tree

## Recently Updated

- [APTrust](#) Jul 15, 2014 • updated by [Scott Turnbull](#) • [view change](#)
- [2013-09-23 APTrust Working Group Meeting](#) Sep 24, 2013 • updated by [Scott Turnbull](#) • [view change](#)
- [2013-09-10 APTrust Working Group Meeting](#) Sep 17, 2013 • updated by [Scott Turnbull](#) • [view change](#)
- [2013-09-17 APTrust Working Group Meeting](#) Sep 17, 2013 • updated by [Scott Turnbull](#) • [view change](#)
- [2013-09-03 APTrust Working Group Meeting](#) Sep 03, 2013 • updated by [Scott Turnbull](#) • [view change](#)
- [2013-08-27 APTrust Working Group Meeting](#) Aug 27, 2013 • updated by [Scott Turnbull](#) • [view change](#)
- [2013-08-20 APTrust Working Group Meeting](#) Aug 20, 2013 • updated by [Andrew Curley](#) • [view change](#)
- [2013-08-20 APTrust Working Group Meeting](#) Aug 20, 2013 • created by [Scott Turnbull](#)
- [2013-08-13 APTrust Working Group Meeting](#) Aug 13, 2013 • updated by [Scott Turnbull](#) • [view change](#)
- [2013-07-30 APTrust Working Group Meeting](#) Aug 13, 2013 • updated by [Scott Turnbull](#) • [view change](#)
- [2013-07-23 APTrust Working Group Meeting](#) Jul 30, 2013 • commented by Anonymous
- [2013-07-23 APTrust Working Group Meeting](#) Jul 25, 2013 • updated by [Scott Turnbull](#) • [view change](#)
- [2013-07-16 APTrust Working Group Meeting](#) Jul 23, 2013 • updated by [Scott Turnbull](#) • [view change](#)
- [2013-07-02 APTrust Working Group Meeting](#) Jul 16, 2013 • updated by [Scott Turnbull](#) • [view change](#)
- [2013-06-25 APTrust Working Group Meeting](#) Jul 02, 2013 • updated by [Scott Turnbull](#) • [view change](#)

[Show More](#) 

## 2 APTTrust Conference Line

---

### 2.1 Phone Instructions

---

Dial-in Number: (805) 399-1200

Participant Code: 900313#

### 2.2 Skype Instructions

---

1. Add freeconferencecallhd.8053991200 as a contact in your skype account:
  - a. copy and paste " freeconferencecallhd.8053991200" (the entire name) into your contact search area
  - b. add as a contact
2. After freeconferencecallhd.8053991200 is listed in your contact list, simply connect to that contact as you would an individual.
  - a. Enter the following code through the Skype Keypad: 900313#
  - b. You must enter the code in the Skype keypad which is located in various areas depending on your operating system and Skype versio

## 3 Institution Codes

These codes are used for identifying individual member institutions in various parts of the APTTrust system. Individual partners have been asked for their feedback on which code to use and these will be updated as we receive them.

Codes should:

- 3-4 characters long.
- Disambiguate the institution.
- Have a high probability of uniqueness.

Institution	Code
Columbia University	
Duke University	
Emory University	
Johns Hopkins University	
University of Maryland	
University of Michigan	
University of North Carolina at Chapel Hill	unc
North Carolina State University	ncsu
University of Notre Dame	
Stanford University	
Syracuse University	
University of Virginia	uva



## 4 APTrust Meeting Notes

---

### 4.1 2012-08-06 - APTrust Developer Call

---

#### 4.1.1 Attendees

Donna Tolson  
Robin Ruggaber  
Mike Durbin  
Michele Kimpton  
Chris Wilper  
Andrew Woods ★  
Andrew Curley  
Tim Sigmon  
Adam Soroka  
Greg Jansen

#### 4.1.2 Discussion

##### Ingest

- Fedora server on UVa is up
  - server connected to provided I.P.
  - Chris W. will be accessing server next
- 100 objects, ~2 GB

##### Content Flow

1. First ingest into DuraCloud
  - ...then into Fedora
2. Content comes into DuraCloud from many different institutions
  - DuraCloud can act as a holding area
  - Each institutions would push to its own space
  - Content can be served as managed directly from DuraCloud
3. Fedora needs to be the gate-keeper of access controls
4. Fedora is needed for institutional policies

5. Decision: ChrisW. / MikeD. will work on syncing test content into both:
  - DuraCloud
  - Fedora
6. Initial test content is Fedora content
  - Later tests will come from non-Fedora content that has FOXML
  - Later tests will include content from multiple institutions
    - Institutional contributions may pass through local storage areas before pushing to APTrust
      - This could impose significant effort on the source institution
    - Or, institutions can push content to an APTrust staging repository
    - Test content size/type is designed to stress limits of system/integration
7. Mike and Greg will need to coordinate ingest flow, and object creation
8. Mike to be in touch in a few days
9. Chris will be in touch with Greg to use CloudSync

## Survey results

- Notre Dame
  - Michigan
  - ...
1. Fair amount of alignment around interest in preservation, versus repository services
  2. A variety of formats/content-types will be submitted
  3. Total data for phase-1: ~40 TB?
  4. Donna to post survey summaries in the wiki

## Next In-person Tech Meeting

- Oct 9th is the next UL meeting
  - Would be nice to have something to show by that date
  - Technical progress should be reported at this meeting
  - Will be clarifying business model at this meeting
- Tentatively setting next f2f tech meeting for Sept 24-25th
- Duraspace will not be attending the Hydra partners meeting

## DPN Chicago Tech Meeting Discussion

## 4.2 2012-08-20 - APTTrust Developer call (by Donna Tolson)

---

### 4.2.1 Attendees

Donna Tolson  
Robin Ruggaber  
Mike Durbin  
Michele Kimpton  
Andrew Woods  
Andrew Curley  
Tim Sigmon  
Adam Soroka  
Greg Raschke

### 4.2.2 Absent

Chris Wilper  
Greg Jansen

### 4.2.3 Discussion

#### Ingest

Ingest tests going well, have tested the following:

- Ingest of UVa Fedora content into DuraCloud
- Ingest of UVa Fedora content into APTTrust Fedora (a few policy errors, but Chris & Andrew C are addressing)

Next plans to test:

- Ingest of large file (<5G) (Mike will work with Chris to get one ready)
- Ingest of non-Fedora content from local file system to DuraCloud and APTTrust Fedora

Q - Does ingesting institution have to log into CloudSync locally? A - All objects & external references need to be open to CloudSync

Discussion: Ingest of non-Fedora content can't currently be started from source. Will try running CloudSync locally. UVA still needs local discussion to determine if all deposited content will first be put in Fedora.

Discussion re: normalization utility: will do 2 things

1. Create APTrust foxml & metadata
2. Split objects into foxml & externally managed datastreams

Q - Do externally-managed items have to be processed by Fedora? A - Entire object must be transferred into cloud before objects can be split up. Once in DuraCloud, should be able to create externally managed content and reference it in Fedora.

[Mike will document specifics of planned ingest tests.](#)

## **Cost models:**

- lifetime preservation model would be attractive to community
- consider asking Kim Thanos to do some research on cost models

## **Survey results:**

- Priorities for Phase II highly varied
- Documentation to become trusted digital repository is enormous; would impact cost. Perhaps we could pursue grant to cover documentation need?

## **Next communication:**

- Robin will set up googlegroup on aptrust.org
- Next conference call is on Sept. 4 to avoid Labor Day. Michele can't join us.
- Next f2f tech meeting set for mid-day Sept. 24 - late afternoon Sept. 25. Robin will work on agenda for development group.
- Duraspace will not be attending the Hydra partners meeting

## 4.3 2012-08-20 - APT Strategy Meeting (by Andrew Woods)

---

### 4.3.1 Attendees

Andrew Curley

Mike Durbin

~~Greg Jansen~~

Michele Kimpton

Greg Raschke

Robin Ruggaber

Donna Tolson

~~Tim Sigmon~~

~~Adam Soroka~~

~~Chris Wilper~~

Andrew Woods ★

### 4.3.2 Call-in Details

Dial-in Number: (805) 399-1200

Participant Code: 900313#

### 4.3.3 Discussion

#### Ingest Testing

- Chris sent current summary
- Need to run big file test
  - 5-GB is a "big file"
  - not this week, as Mike and ChrisW are both on vacation
- Four tests
  1. fedora to cloud - done w/ UVa test content
  2. fedora to aggr - done w/ UVa test content
  3. local to cloud - not yet done
  4. local to aggr - not yet done

- Will need to add on to 1. and 3. to test moving content from cloud to aggr

- Will begin testing UNC content next week when Greg J is back
- Question: does cloudsyntax currently support external references to duracloud? Answer: ??
- Question: do externally-managed items have to be processed by Fedora?  
Answer: entire object must be ingested into cloud before externally-managed content is identified, referenced in Fedora
- Will need to work through issues of local policies needing to be available to CloudSync
- Need to document tests
- Need to document workflow of system
  - Add topic to next f2f meeting
- Currently in the proof-of-concept phase
  - At the end of this we will know what works and what new code needs to be developed
- We have not eliminated any options
  - Institutions to aggr-repo?
  - Institutions directly to duracloud?

## Content Survey

- Versioning received highest rating around need in APTTrust
- Need to quickly come up with a cost structure
  - Preservation is an on-going cost, will likely get push-back from users
  - Explore possibilities for a single, lifetime cost
  - Consider HathiTrust model
- Advisory committee should be working on the business model
  - Kim Thanos could be a resource in developing a model
- There is a fairly wide spread of user needs/requests
  - May need to distinguish items on survey that are likely to happen, and those that are more tricky to implement
  - There is high interest in APTTrust being TRAC certified - not difficult, but will involve a lot of documentation and expense - perhaps pursue grant?
- Need to develop a Q&A for content/technical user-base
  - webinar?
- Advisory group call this Friday
  - next advisory f2f is Oct 9th

## Next Steps

- Next call: Tuesday 4th @1pm
- Ingest tests will be ongoing
- Mike: tests will be documented in wiki
- Robin: to share arch docs
- Robin: create a technical APTTrust google-group?
- Greg: working with advisory group

- Mike/Wilper: running local cloudsync

## 4.4 Attendees

---

Donna Tolson  
Mike Durbin  
Chris Wilper  
Andrew Curley  
Tim Sigmon  
Adam Soroka  
Greg Jansen  
Bradley Daigle

### 4.4.1 Absent

Robin Ruggaber  
Michele Kimpton  
Andrew Woods

### 4.4.2 Discussion

#### Ingest testing

Chris W. completed tests from UVa staging to DuraCloud and Fedora. Had just completed same test with UNC content. Basic access worked fine, but socket timeouts occurred between services. Can be fixed by adjusting CloudSync defaults.

Still need to test large file - expect some problems ingesting to both DuraCloud and Fedora.

Ingested content took a bit longer going into DuraCloud than into Fedora. Still need both environments, to use check sum functionality of DuraCloud (and maybe multi-tenancy once multiple institutions hitting the system at once) and edit & management functionality in Fedora.

Discussion of nearline vs. external storage. Could separate objects into DuraCloud & Fedora as discussed on 8-20; CloudSync could be adapted to do this.

While several options were presented, before beginning tests that require changes to the setup of fedora (ie, pointing to different storage) the tests and work that need to be completed regardless of the storage were to be worked on first. (specifically the large file upload issues)

The solution of putting Fedora "directly" on top of DuraCloud (where fedora managed storage is accessed through DuraCloud) will require resumed development on an Akubra plug-in for DuraCloud storage. The option where large content in DuraCloud is left in DuraCloud and linked to via an external reference datastream in Fedora requires changes to CloudSync. Since both require development before testing could begin, and it was not 100% clear to all participants whether both (or either) would meet all the business requirements for APTTrust, Mike volunteered to describe and place in the wiki the several options and their implications as far as he could tell.

## Performance issues

Robin had emailed question to group, but no one was certain what she meant. Donna will ask her to follow up to group.

## Cost issues

As long as Fedora & DuraCloud are in same availability zone, no additional cost for transferring content between the two. If not in same availability zone, additional costs would incur.

## Next communication

Conference call on Sept. 17. Sept. 24-25 is next face-to-face meeting in Charlottesville.

# 4.5 2012-09-24 & 25 - APTTrust Face-to-Face Meeting - Charlottesville

---

## 4.5.1

- [Attendees](#)
- [Location and Time](#)
- [Agenda](#)
  - [Monday](#)
  - [Tuesday](#)
- [Architecture](#)
- [Whiteboard](#)



- Discussion
  - Summary of Meeting Accomplishments
  - Detailed Accomplishments
    - Determined what to show and tell for Steering Committee on October 9th
    - Development tasks assigned
    - Basic requirements for administrative interface outlined
    - Milestones for December 2012
  - Day 1
    - Attendees:
    - APTrust Expectations
    - DPN Update
    - Review Ingest Workflow Scenarios
    - Journaling/Logging
  - Day 2
    - Administrative Use Cases
    - What can we get from current APIs? What should each application provide?
    - Discuss Demo Needs
    - Goals for the end of the year 2012:
    - What are we obligated to DPN in the next six month?
    - Work to be done

## 4.5.2 Attendees

Donna Tolson  
Mike Durbin  
Andrew Curley  
Tim Sigmon  
Adam Soroka  
Greg Jansen  
Bradley Daigle  
Robin Ruggaber  
Martha Sites  
Michele Kimpton  
Andrew Woods

## 4.5.3 Location and Time

Brown Science and Engineering Library, Room 148 (Clarke Hall)  
Monday (9/24): 12pm - 5pm  
Tuesday (9/25): 9am - 3pm

## 4.5.4 Agenda

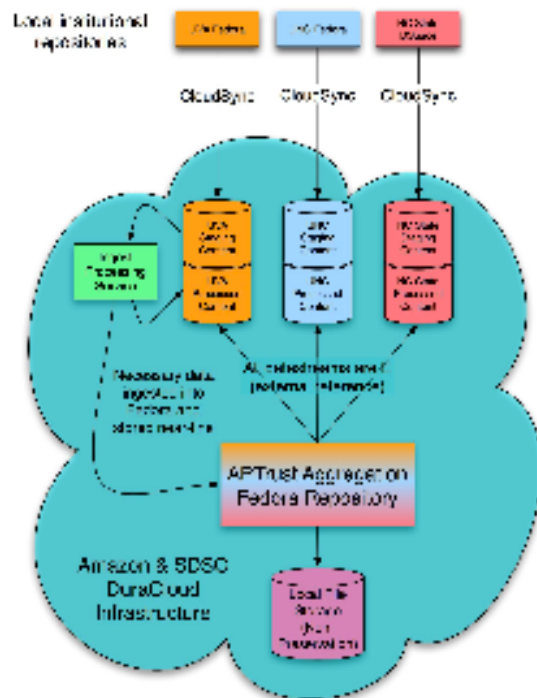
### Monday

1. *Noon – 1:00*: lunch and agenda review
2. *1:00-1:15*: APTTrust Expectations – Martha
3. *1:15-1:30*: DPN Update – Andrew, Adam and Tim
4. *1:30-3:30*: Review Ingest Workflow Test Scenarios(4) & Results - Mike Durbin, Andrew Curley and Greg Jansen
  - a. How content moves through the system
  - b. Description of metadata that is generated and passed along the flow
  - c. What processing happens at each point
  - d. Human interactions
  - e. Alignment on Ingest Approach
5. *3:30-3:45*: Break
6. *3:45-4:15*: Journaling/Logging Needs
7. *4:15-5:00*: Clarify Work to be Done - (cloud process utility, Cloudsync, Ingest Utility, etc) - team

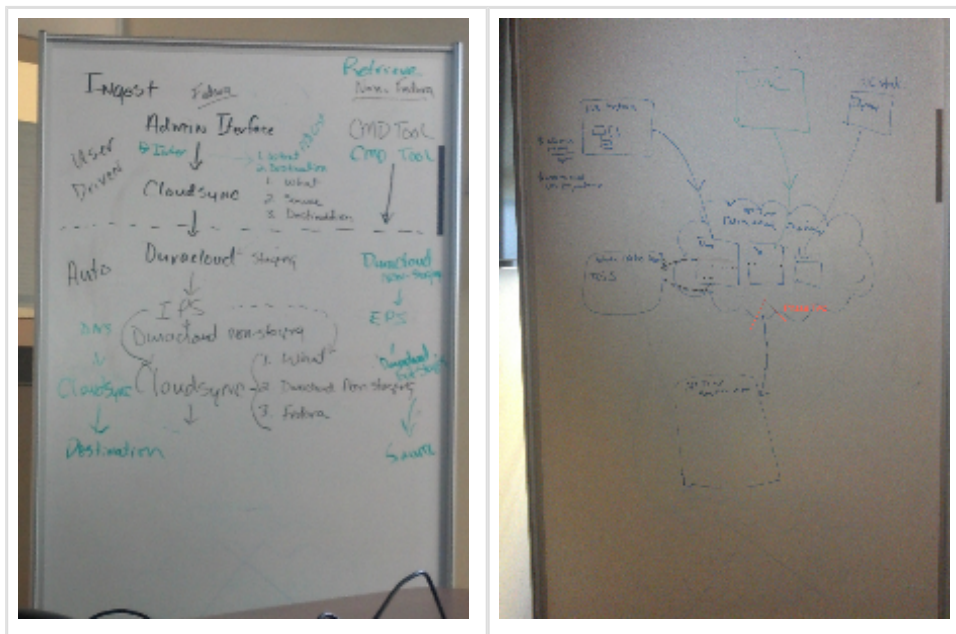
### Tuesday

1. *9:00-9:30*: DPN Planning - Andrew, Adam, Tim
2. *9:30-10:00*: Administrative Use Cases - Interface - Bradley
3. *10:00-11:00*: Identify Administrative Interface Requirements (Cloudsync, Duracloud & Fedora) - Adam, Andrew Curley, Andrew Woods, Mike Durbin
4. *11:00-11:30*: Discuss Demo Needs/Possibilities for Oct 9 - Donna
5. *11:30-noon*: Existing Milestone Review - Donna
6. *Noon – 1:00*: Lunch
7. *1:00 – 2:00*: Identification for Remaining Work - Donna
8. *2:00 - 2:30*: Finalize Roadmap & Assign Milestone Ownership – Donna
9. *2:30 – 3:00*: Next Steps - Martha

## 4.5.5 Architecture



## 4.5.6 Whiteboard



Ingest Workflow

Top-level Architecture

## 4.5.7 Discussion

### Summary of Meeting Accomplishments


1. Overall architecture decided.
2. Determined what to show Steering Committee on October 9th.
3. Development tasks assigned.
4. Basic requirements for Administrative Interfaced outlined.
5. Milestones for December 2012 set.


### Detailed Accomplishments

#### Determined what to show and tell for Steering Committee on October 9th

1. Have Fedora content from two institutions prepared!
2. Working on accomodating large files (i.e. over 5GB) to allow for datasets and A/V material.
3. Show DuraCloud UI for Fedora content in preservation environment.
4. Assignment of tasks (as outlined below).
5. APTTrust architecture will not need refactoring to accommodate DPN milestones.

#### Development tasks assigned

1. Cloudsync
  - a. Chunking - Andrew W
    - i. Test functionality with large files - UVA and UNC?
  - b. Checksumming prior to transfer - Andrew W
  - c. Test ingest from DuraCloud to APTTrust Fedora - Andrew W (get some content from UVA/UNC)
2. Ingest Process Service
  - a.  Write skeleton that copies file from staging space to new space - Andrew W
    - i. Business logic inside of this - Mike
    - ii. Super intelligent produced PID list for transport OR in CloudSync OR Admin tool - Mike
3. Create logging of ingest workflow steps (needs to be fleshed out more) - Mike
4. Documentation
  - a. Architecture diagram updates
    - i. Overview (synthesis for non-tech people) - Donna
  - b. Detailed - OLE (Adam for DPN, Robin, Mike), DuraSpace for tools

5. DSpace Ingest
  - a. NC State contact - Mike will contact Marcus.
  - b.  DuraCloud for Research wrapper and analogous to command line tool for DSpace packaging - DuraSpace (Andrew W)
6. Administrative Interface
  - a. Main work on web app - DuraSpace (Danny and Andrew W)
  - b. Wireframes - UVA (J. Boggs?)
  - c. Solr - Mike
  - d. Export functionality - Mike
7. Client-side tools
  - a. Adapting command line tool for non-Fedora based ingest and validation of metadata - Mike
  - b. Export functionality - Mike

## Basic requirements for administrative interface outlined

The admin interface should provide the facility to:

1. Show the status of current CloudSync operations.
2. Prevent two CloudSync operations from one partner institution from occurring simultaneously.
3. Provide some user authentication with simple permissions. (Shibboleth enabled eventually)
4. Allow a user to search/browse their APTrust holdings
5. Allow for on-demand fixity checking.
6. Provide audit reports.

The admin interface should answer the following questions:

1. What's in there?
2. What did we put in in the last X [timeframe]?
3. Is X in there?
4. What is everything that is in APTrust for my institution?
5. Why didn't this make it in?
6. How many things are in there?
7. What is the status of this ingest?
8. How much is the bill going to be this [timeframe]?

## Milestones for December 2012

1. DSpace content can preliminary be put through Ingestion Processing Service
2. Implement Fedora architecture
3. Make modifications to CloudSync
  - a. Chunking
  - b. Checksum prior to transfer

4. New ingest service module
  - a. APTTrust normalization utility (now the IPS) is moved into the cloud.
  - b. Local institutions don't need to worry about how to package their content locally for APTTrust.  
Removed local institution staging, saves money and time. Leverage power of the cloud!
5. Administrative interface
  - a. List of features and stories above.

## Day 1

### Attendees:

Andrew Curley  
Bradley Daigle  
Mike Durbin  
Greg Jansen  
Michele Kimpton  
Robin Ruggaber  
Tim Sigmon  
Martha Sites  
Adam Soroka  
Donna Tolson  
Andrew Woods

**All:** General introductions.

Agenda Setting and Amendments to Current Agenda

**Michele:** Need to know what work specifically needs to be done on CloudSync since Chris left. Andrew W and Chris reviewed code before he left.

**Donna:** I want to make sure that I capture deliverables.

## APTrust Expectations

**Martha:** When we last met we believed that by October - December we would have the ability to demonstrate ingest and workflow. For my own expectations, I would like to demo something on October 9th at the partner steering meeting. I would like to demo something, not fancy, and show that we do have ingest capability. Want to predict our timeline and state our milestones. Want to talk about blockers or stoppers. Want to say what it will take actually to go into the new year with ingested content.

**Mike:** Uncertain what is doable. Current process isn't showy. What type of demonstration is appropriate?

**Martha:** Just a steering committee so it doesn't need to be showy. I want to be able to answer people questions about whether content is there. Also want diagrams of the workflow.

**Robin:** DuraSpace's demo back in January was powerful and a testament to the usefulness of demonstrations.

**Martha:** As long as we can keep these things in mind and know that this is about participation, we will be able to demonstrate. This whole team needs to speak freely as to what it is. Speaking with one voice about the architecture we have chosen and how we are going to move forward. Especially wants to know what the blockers are so she can help move them.

**Greg:** We should add for near term issues of scalability and throughput. UNC can put a lot of content in but how much will we really be able to achieve. Will need performance tests. Most useful to do analysis on content side.

**Adam:** We are not entirely sure as to what ingest means in terms of technical apparatus.

**Martha:** Given survey results, we may need to identify a few institutions who are willing to make their content available. Not practical to get this kind of information from lots of partners. Same concerns for the cost model. We don't know how to price this initiative without knowing how much content there will be.

**Michele:** Just getting the content to transfer over https protocol was not easy so there will be practical issues of moving stuff that will undoubtedly need to be addressed. Perhaps each institution will have their own unique problems related to the transfer of content.

**Martha:** We have assumed that someone with technical capabilities will work with each institution because there will be issues peculiar to each institution.

**Tim:** We obviously want to use Internet2 and hopefully Amazon will come online with Internet2.

**Donna:** There may be a policy implications about allowing access, or at least bulk access, to production infrastructure. We had such experiences here at UVA.

**Robin:** Are we taking from production repo or staging repo or a staging disk? Unknown times when content will be downloaded caused sysadmins alarm.

**Adam:** We should take into consideration that we may need to have trucks with tapes (a more practical approach) rather than imposing technical needs.

## DPN Update

Steven Morales is the program director for DPN. Coordinating various sub-groups.

**Andrew W:** Just attended our second face-to-face. In between meetings there was a weekly call. Technical team convening. Making progress with defining what the responsibilities are for each of the five DPN nodes (APTrust, Chronopolis, HathiTrust, UT and Stanford). Tech team is comprised of each of the five nodes plus DuraSpace. Documented on the same wiki as APTrust. Consensus points around somethings. General milestones like use of centralized Github account. The general idea is that there will be two communications channels between DPN nodes: messaging for control (i.e. I have content to ingest, like JMS) and content transfer channel, starting implementation with rsync over https. In October messaging prototype (i.e. standing up messaging framework for each node) will have generated packages or spectra of packages with using Bagit. Method of transfer will be bags. In November/December timeframe DPN nodes will be prototyping transfer of content (nothing big). At that point, there is believed to be enough to warrant a face-to-face at CNI.

**Adam:** There has been discussion about technical architecture, what is to be implemented. Mutually agreed upon need to "brighten" material. Consensus is that there needs to be more definition around requirements. Versioning is also another problem.

**Martha:** So we see no conflicts between us and DPN.

**Adam:** By checking the DPN box, there will be a lot of changes for the user in APTrust. Different admin interface, different actions. And the choices may be irrevocable. There will not be an additional DPN interface provided by APTrust. Brightening will be a message coming from another node.

**Andrew W:** There will probably be no software that is DPN, just a common wire protocol between the nodes.



## Review Ingest Workflow Scenarios

**Mike:** Reviewing architecture. Look at picture. Discussion of APTrust objects. Not adding metadata to objects, just linking to existing objects.

**Adam:** Cloudsync can take an arbitrary SPARQL query to determine objects.

**Mike:** We might need to tinker with admin interface to ensure that the queries to determine content are run at appropriate times. The APTrust utility just takes the entire contents of a Fedora repo and creates APTrust objects. (This is now overwritten).

**Adam:** One needs to be concerned about over eager SPARQL query which will shut down Mulgara.

**Mike:** There were actually two routes. Fedora to Fedora and Fedora to DuraCloud. There should be no blockers to working on chunking content over 5GB. The second goal of the ingest procedure is to use CloudSync from DuraCloud storage to aggregation repository. The aggregation repo is using EBS (Elastic Book Store, the main type of storage Amazon provides for their compute service) not S3 (Simple Storage Solution, the vastly more reliable cloud storage service). The first use of CloudSync is using S3. If we ingest directly to Fedora, we need to transfer the content off EBS to S3. May need to modify Akubra.

**Michele:** SDSC will be out most likely near term partner but they are not proximate geographically with APTrust's initial partners who are predominantly on the east coast.

**Mike:** Not relevant to architecture now but may be important when assessing performance metrics.

**Robin:** We may not want to bank any development on making Akubra work since it might be going away in the next 9 months, according to Fedora Futures meeting.

**All:** We are not going to do a multiple ingestion workflow where everything gets ingested to both S3 and EC2 because it would have to be a super CloudSync and Robin doesn't want to invest lots of development time.

We have decided that we will not be using Akubra to store Fedora objects and datastreams in S3. Based on research that Chris and Andrew W did, the group has decided to use CloudSync to put data and objects into S3 and then copy pertinent data to EC2 while externally referencing other content back to S3. Presumption is that EC2 instance and EBS will die so we need to rebuild.

**Mike:** We need to worry about new versions because since Fedora is being updated from S3, there is no intrinsic notion of Fedora versioning. Now we need to pick apart this decision. What to do with new versions? CloudSync will have to check if object has new version. CloudSync will never have to overwrite. Will need to change CloudSync to not move everything every time. Need to worry about things that are DPN bound from moment one. We cannot have our utility overwrite objects and datastreams in DuraCloud that are flagged as DPN.

**Greg:** Question about file path.

**Adam:** DPN nodes are going to be at least as versioned as any repository with content coming into it.

**Mike:** CloudSync will have to be initiated from an interface that allows selective syncs and throw errors when trying to delete DPN. Create another rules layer that then allows for development outside of CloudSync and allows CloudSync to be the tool this is now rather than something greater than it should be.

**Adam:** Not crazy to ask institutions to not tinker with repo during transfer time. Let's put journaling on the table for later. Need to consider doing eventing within the OSGI container.

**Mike:** Need to write new service to intercept the copy.

**Andrew W:** I will write a blank method in the chain and we can enter our own code.

## BREAK

**Mike:** We now have a placeholder in DuraCloud to store all of our work. So we got that going for us.

**Andrew W:** Three stores CloudSync is aware of: Fedora, Duracloud and directory.

**Mike:** For now we should purge everything in the aggregation repo (as it currently exists) and repopulate it from DuraCloud. For non-Fedora content, we now have the option to normalize the content in the cloud rather than before it enters the cloud. Perhaps we should experiment with a DSpace institution and use our new service to transform it into more Fedora looking instance while in the cloud. With DSpace sourced content, we need to store their original content or provide a lossless translation so that they can retrieve their APTrust object as a DSpace object, not Fedora. Rather than rewriting the reverse translation we should just store the original and then worry about that reverse translation later. Wait until phase 2? New system will be called the Ingest Processing Service (IPS). Initially there will be one service but eventually there will be n number of services so as to provide a decoupled workflow.

**Adam:** Perhaps we need to have institutions provide a URL where they indicate where the rich metadata is.

**Mike:** That might be unnecessary because we can have them provide that within Fedora.

**Adam:** I don't know if that is possible.

**Michele:** Where is the fixity number coming from?

**Mike:** Don't know. This might be a reasonable addition to CloudSync.

**Andrew W:** I think there is already a JIRA ticket for this.

**Donna:** Bradley can you make sure that the content people are ok with only getting back what they put in and nothing else?

**Mike:** We might make different rules for the "aptrust" object.

**Greg:** How do we handle the aptrust set? What are the rules for a set?

**Mike:** There are few rules.

**Adam:** For DPN bound objects there might be some more complicated enforcement because it is unclear whether this is allowed. Also, it is smart that we are leveraging the institution's own repository. But we are assuming that people are operating rationally and we must accommodate their stupidity.

## RULES:

- Cannot delete "DPN" content
- Can replace or provide a new version.
- Valid APTrust metadata must accompany all submissions.
- Logging is stored with objects but what partner retrieves will not have that content unless requested.

## Journaling/Logging

**Adam:** There has been a distinct lack of discussion on this. Surprised. Need workflow logging, storage logging.

**Robin:** Do we need anything more from DuraCloud to satisfy our needs?

**Mike:** We will have to organize the data referencing the data they are familiar with, not the post-transformation objects. Most pertinent to non-Fedora content (i.e. DSpace).

**Adam:** OMG, let's not have to manage id matching. With Fedora it would be less of an annoyance.

**Mike:** Will have to repackage the content.

**Adam:** Assembling what is in the object from Fedora and taking information from DuraCloud and presenting that information to the user in the nice report. Storing the former name and the new name will be part of the Fedora object. When they retrieve the object this logging information is not delivered back to the user. The log is retained in the aggregation repo within the object and is retrievable through the admin interface. We need not expose md5 numbers to the partners without violating the preservation needs.

**Michele:** Partners do want that information though.

**Adam:** But we are not going to retain the checksum before transfer because that is done on partner hardware. Perhaps we can create a service to rerun checksums on Fedora objects and exclude certain datastreams in order to provide the checksums against the original objects.

**Mike and Andrew W:** Concern over reversibility of objects. They may need to make a decision about whether the object crosses a threshold.

**Donna:** Isn't this a problem with DPN content? Can we provide checksums for their content.

**All:** We are never going to do anything with DPN content that would require that. We are agnostic about that content and have a low level of services for that content.

**Mike:** We are never going to get TRAC certification if we cannot reproduce exactly what was contributed to us by the institution but instead only provide our post-transformation content. If we are taking content that is in compressed format, we may need to retain two copies of all their content so we can ensure that the content is exactly the same.

**Donna:** Although the survey results indicated that the TRAC is desirable, it may be too expensive.

**Andrew W:** There is a desire in DPN for all nodes to be TRAC certified because two of the partner already are.

**Mike:** We can reject files based on format. I think it is genuine to reject files based on format (i.e. .zip files).

## Day 2

### Administrative Use Cases

**Bradley:** No formal use cases for administration. What are the options? Would like to have a technical frame of what it will have before he talks to content people so as to guide the conversations.

1. Management of files (single, batch upload/download (configure CloudSync to pull from Fedora), delete.)
2. Reporting

3. Search/browse. This will only be searching the minimal metadata provided for each object for the time being. This will be searching Fedora but Adam warns against using the Fedora admin utility because it has known problems. Do we therefore need to throw Solr in front of this to meet minimum requirements? Are we going to use Blacklight as the admin interface from the beginning?
  - a. **Greg:** Faceting by set would be useful.
  - b. **Michele:** Danny from DuraSpace would be the best candidate for development and is most comfortable using Java. Has previously written the DuraCloud admin interface. The first admin interface could be a throwaway after getting feedback. He also has experience with Solr as well.

Other's suggestions for what admin interface needs to provide:

1. Status of this ingest.
2. How much is the bill going to be?
3. Prevent two CloudSyncs from being run simultaneously from one partner. Show status of current syncs?
4. Permissions for logging in? Single user? Multi-user with different privileges?
  - a. Robin wants shibboleth enabled. DuraCloud is currently not shibboleth enabled and cannot get it done by January. We will not get certified for TRAC by collapsing identities, but TRAC is not a goal for phase one...
  - b. Logging of which user did what action will have to be recorded in the admin app.
  - c. Currently DuraCloud logs user actions but they are not organized nicely and there is currently no facility to present to the user nicely.
  - d. We will need to bring together logs from various sources to provide the information we need.

Not clear who the agents are who will be using this interface.

Andrew C's stories that must be answered by the admin application:

1. What's in there?
2. What did we put in in the last X [timeframe]?
3. Is this in there?
4. Output manifest
5. Why didn't this make it in?
6. How many things are in there?

Andrew W wants to know what the admin console is as he was not part of these original conversations. It is an application layered over both the Fedora and DuraCloud API.

**Robin:** What should we do about the search mechanism. Will this include the more sophisticated Blacklight /Hydra app? (Group decides that the phase one admin tool should do all the above).

- Robin's comments supplied via email during the meeting: APTTrust will have two interfaces, one for admins which incorporates a subset of admin functions from CloudSync, DuraCloud and Fedora and the second for other people who will need to authenticate with shibboleth and then be able to search and discover. The admin interface will give the admin control to allow access to their site or not, give access to public or not.

Need better link between local filesystem and remote CloudSync. Client side code? DuraCloud Sync tool already does this. Will require local install. We need some pre-validation before we push the files. Is the minimum metadata included? Too much speculation as to what the tool will need. Mike calls "tangent" and group moves on...

Greg suggests Curator's Workbench to add some metadata for filesystem stuff. Creates METS, adds PIDs, does checksums, export Bagit.

## What can we get from current APIs? What should each application provide?

Fedora:

1. Query to determine all the APTTrust metadata objects and then determine if they are all ingested. Have to be able to filter by institution. We do want to use Solr eventually, so should we integrate it sooner rather than later? We have decided that we are using Solr.
2. We are not going to use the admin functions of Fedora.
3. Admin tool will not be changing anything in Fedora since all datastreams are external references.
4. **(Need to put off all delete actions until after December 2012 and decisions have not been made as to where the delete command will start the delete process.)** There will be a need to leave a tombstone in Fedora which Adam says is trivial. We should leave the object in Fedora but purge the content from storage because people will not want to pay forever for stuff they don't want. We would still maintain the APTTrust metadata.
5. Editing is out of scope, so we need worry about that.
6. Phase 2 would be more interesting if RDF is the backbone rather than Solr.

DuraCloud:

1. Given the small number of users and the difficulty in putting shibboleth up before December 2012, individual user IDs may be sufficient.
2. Shibboleth integration by December 2013 is necessary. Use flat file to authorize. Still need to register with DuraCloud app so as to create rules within APTTrust.
3. Fixity checking. Use current functionality (scheduled runs). Need to output reports to admin interface. Merge with Fedora logs with some intelligent processing?
4. Triggered auditing. Value of on demand seems overkill. Can pick lists of objects though. Great selling point but no one wants to do it.
5. Expose reports but don't allow configuration through admin interface.
6. Prevent user from deleting, touching objects.
7. Be able to pull content based on PIDs...?

CloudSync:

1. Start reverse sync back to local institution.

Revising how to deal with zipped DSpace-sourced content. When syncing back to a local institution, do we repack the content or do we stored two copies (normalized/Fedora-ized and original).

*(Insert ingest workflow diagram Robin drew.)*

**Greg:** Need to deal with the complexity of the semantics of syncing.

Mike is doing the validation and working on the skeleton for IPS (all processing services).

## Discuss Demo Needs

Have Fedora content from two institutions prepared! File types are not important, filesize is. No chunking capability currently implemented so nothing greater than 5G has been pushed.

## Goals for the end of the year 2012:

1. DSpace content. (Marcus is our point of contact at NC State)
2. Improved Fedora ingest workflow.
  - a. Ingest workflow that was used during testing has now been replaced, at least on paper, by the work of this meeting
3. Preliminary DSpace workflow.
4. Make modifications to CloudSync
  - a. Chunking
  - b. Checksum prior to transfer
5. New ingest service module
  - a. APTTrust normalization utility (now the IPS) is moved into the cloud.
  - b. Local institutions don't need to worry about how to package their content locally for APTTrust. Removed local institution staging, saves money and time. Leverage power of the cloud!
6. Administrative interface
  - a. List of features provided by above.
7. Future Functionality
  - a. Delete
  - b. Shib
  - c. Export/Retrieval

Want to confer upon the ULs some information about costs. Michele would like to price out via terabyte, compute cycles and human resources. With one copy, \$1300, two copies \$2100 per terabyte per year. These are the APTTrust costs, but how would we charge DPN customers. Transfer up is free, download has a cost but it should be so infrequently done that we should not charge the users. Bradley objects. Describes how we use Stornext and pull content down frequently from that archive. Warns of potential abuse by institutions. Depends on how we are going to use and sell the service. We should shut down abusive users. Michele thinks we should just say bandwidth included. Once we move to Internet2 it won't be a problem. Perhaps we should use Glacier as the APTTrust second copy to reduce costs down to \$800 at SDSC + \$130 on Amazon Glacier.

## What are we obligated to DPN in the next six month?

1. They have separate milestones.

2. Exchange messages and content. Does not necessitate APTrust to get hardware and software up.
3. APTrust is the only node of the five DPN nodes that doesn't yet exist. Other nodes will require some work to be ready for DPN, but they are extant projects with infrastructure.
4. APTrust should be able to meet DPN timelines.

## Work to be done

1. Cloudsync
  - a. Chunking - Andrew W
    - i. Test functionality with large files - UVA and UNC?
  - b. Checksumming prior to transfer - Andrew W
  - c. Test ingest from DuraCloud to APTrust Fedora - Andrew W (get some content from UVA/UNC)
2. Ingest Process Service
  - a. Write skeleton that copies file from staging space to new space - Andrew W
    - i. Business logic inside of this - Mike
    - ii. Super intelligent produced PID list for transport OR in CloudSync OR Admin tool - Mike
  - b. Create logging of ingest workflow steps (needs to be fleshed out more) - Mike
3. Documentation
  - a. Architecture diagram updates
    - i. Overview (synthesis for non-tech people) - Donna
    - ii. Detailed - OLE (Adam for DPN, Robin, Mike), DuraSpace for tools
4. DSpace Ingest
  - a. NC State contact - Mike will contact Marcus.
  - b. DuraCloud for Research wrapper and analogous to command line tool for DSpace packaging - DuraSpace (Andrew W)
5. Administrative Interface
  - a. Main work on web app - DuraSpace (Danny and Andrew W)
  - b. Wireframes - UVA (J. Boggs?)
  - c. Solr - Mike
  - d. Export functionality - Mike
6. Client-side tools
  - a. Adapting command line tool for non-Fedora based ingest and validation of metadata - Mike
  - b. Export functionality - Mike

## 4.6 2012-10-01 - APT Strategy Meeting

---

### 4.6.1 Attendees

Andrew Curley

~~Mike Durbin~~

~~Greg Jansen~~

Michele Kimpton

Greg Raschke

Robin Ruggaber

Donna Tolson

Tim Sigmon

Adam Soroka

Bradley Daigle

Andrew Woods

### 4.6.2 Call-in Details

Dial-in Number: (805) 399-1200

Participant Code: 900313#

### 4.6.3 Agenda

1. Review actions and status from [face-2-face meeting](#)
2. Coming up with more detailed requirements for the web application
3. UI design for web application- who will do this?

### 4.6.4 Discussion

#### Discussion of Andrew Curley's f2f meeting notes

1. Andrew W. to transcribe PDF notes up on the wiki
2. In page1 diagram:
  - "Non-preservation storage" to be changed to "local filesystem (non-preservation storage)"
  - Discussion around reconciling "staging" content and "processed" content cylinders
3. Donna to work to update first page architecture for next week's presentation



## Work-to-be-done

- Need to layer these tasks against a timeline

## New hire

- Scott Turmbull
- Will be coming on-board in Nov
- Lead technical role for APTrust
- Will help to fill two new positions
  - DPN and aggr repo
  - Workflow vis-a-vis aggr repo
- Likely will be participating in next year's implementation
- Likely will be helping define/design phase-2 effort

## DSpace content

- Mike D. has received some DSpace AIPs
- In the production flow, these will likely come straight from DSpace

## Admin Console Requirements

- UVa to come up with wireframe designer/developer
- Need to develop detailed requirements for webapp
- Need to define a user community to help define use
  - Would like to have a non-fedora content creator involved
- ARL Spring meeting marks a major milestone
- Content ingest sources
  - fedora
  - dspace
  - filesystem
- Phase-2 admin requirements may vastly expand the current scope
- Add requirements to wiki
- Schedule technical calls
- Be thinking about repository functions that will likely be needed in future phase

## 4.6.5 Actions

## 4.7 2012-10-09 - APT Admin Console Meeting

---

### 4.7.1 Attendees

Danny Bernstein

Andrew Curley

~~Bradley Daigle~~

Mike Durbin

~~Greg Jansen~~

Andrew Woods

### 4.7.2 Call-in Details

Dial-in Number: (805) 399-1200

Participant Code: 900313#

### 4.7.3 Agenda

1. Set context, review status / timeline / etc
2. Review initial [functionality and mock-up](#)
3. Define backend support data
4. Define detailed requirements

### 4.7.4 Discussion

#### Content Ingest

1. Goal to avoid client-side content
2. Ingest reporting at the level of an ingest request
  - Initially minimize new code by using the data we have
3. Timeline view of provenance events could be useful per item
4. Req: prevent conflicting ingest operations
  - e.g. overwriting a DPN-bound object

## Content Discovery

1. Solr work is underway
2. Pid search is required
3. Looking at possibly using GSearch
4. In terms of messages and ingest status, may want a stream of status messages

## Reporting

1. Checksum report
2. Storage logs/graphs lifted up from DuraCloud
3. Provenance log is merge of processes and tools that touch content
4. Should go through surveys to find other reporting requirements
5. Clarify DPN reporting needs
6. Billing
  - Maybe split out

## Export

1. Need to discuss what are the reqs for the first round?

## Security

1. Do not need to consider acct creation early on
2. User can only see content from their institution
3. Need a super user

## Terminology

1. Before copy: file
2. After parsing: APT package
  - everything that falls under an APT metadata object

## 4.7.5 Actions

Description	Owner	Status

## 4.8 2012-10-15 - APT Strategy Meeting

---

### 4.8.1 Attendees

Andrew Curley  
Mike Durbin ★  
Greg Jansen  
Michele Kimpton  
Greg Raschke  
Robin Ruggaber  
Donna Tolson  
Tim Sigmon  
Adam Soroka  
Bradley Daigle  
Andrew Woods

### 4.8.2 Call-in Details

Dial-in Number: (805) 399-1200  
Participant Code: 900313#

### 4.8.3 Agenda

1. Partners' meeting update
2. Web admin console plans and status
  - a. Jeremy Boggs?

### 4.8.4 Status

1. Web admin console had the first [meeting](#) last week of what will be a standing weekly meeting (@3pm ET on Tues)
  - a. Requirements and early mock-up are being collected [here](#)
  - b. Expectation is that after next week we will have enough detail to begin handing off work to Jeremy Boggs
  - c. Depending on team availability, we will then push for an initial implementation as early as possible

## 4.8.5 Discussion

Present: Robin, Mike, Andrew W, Tim, Adam, Andrew C, Greg Jansen

Andrew asked about experience with mocking tools. Neither Mike or Tim had any experience.

No one who was present at the partners meeting was on the call, though second-hand we heard that Donna said it went really well.

- Jeremy Boggs confirmed that he could begin working by October 30th,
- in an iterative approach using 2-4 weeks.
- Robin made clear that he could produce HTML and CSS
- Would it be useful to have him on earlier calls, Bethany (his boss) recommended that he not join until the design was more finalized.
  - Andrew W and Mike concurred
  - Andrew W would not be at the Tuesday call in a week, so its our goal to get the design worked out

Andrew W gave an update about the UI design meetings.

- Danny Bernstein will be doing a lot of the work and was able to make the first meeting
- the meeting notes are available on the wiki

Skype dropped many users from the call at 3:11PM.

Greg gave an overview of the partners meeting:

- discussed architecture, everyone was satisfied and didn't have many questions
- got a finances update from Martha
  - all funded through the end of 2013
- business model will work better if more partners are found
- next step is to have a conference call with content liaisons (by early November)
  - are the questions that should be asked of them?
- Donna and Martha will be doing site visits for everyone by Mid-January to firm up survey responses
  - Robin reported that there were some self-contradictory responses to original surveys
  - Greg confirmed this and indicated we should take the results with a grain of salt
- There was a staffing update, but we're aware that Scott will begin on Nov 27
- There will be two more hires from the one posting

Regarding the business model:

- the more partners the better (broader base) "maybe we can double by next year", or "maybe we can double every year"

- the leaner the organization is at the top the better development environment
  - other organizations can grow their development team, but that increases cost
  - instead AP Trust could focus on facilitating involvement from partners

Spruce Project in the UK could be an example of a similar sustainable business model.

Greg could not come up with any questions or comments about the technical architecture that were brought up.

Robin reported and Donna said people were surprised that we were taking in video content and other types of content.

Greg reported that when people saw the DuraCloud pie chart with mime types, that people liked the ideas that you could find other people who had similar content (potentially two people could collaborate over solutions involving rare or obsolete formats).

Robin asked if there could be an interface where people could see across institutions (in a presumably read-only way). Andrew W responded that there had been discussion in the past, but that there's the natural obstacle of content control concerns, and no current push for that feature. Likely this would be handled at a higher level in AP Trust (possibly the Admin interface, or another application) rather than any going directly to DuraCloud.

Possible ways to engage content liaison group:

- use them for feedback on UI design for Admin Interface
  - might have to weigh contradictory opinions, but would be worthwhile
- possibly use Google Hangouts

Adam asked about phase 2 desires. Greg responded that the goal of finding common ground, specifically regarding metadata standards or structures was expressed. The goal is that once there's a big pot of metadata, one can build services on it. Everyone knows its complex.

Final report from the meeting is that there is a plan to go forward with a program session at ARL, and have a partner meeting at ARL.

AP Trust will have presentations at CNI conference in December. Robin will be putting in a proposal for a panel discussion.

When asked about the DPN effort, it was reported that it's going "fairly well" (Adam) and has taken a "turn towards the technical" (Andrew W) to do initial prototyping and less "theoretical and controversial questions".

## 4.8.6 Actions

Andrew W asked everyone to review the [AP Trust administrative interface](#) web page and add/update the content.

## 4.9 2012-10-16 - APT Admin Console Meeting

---

### 4.9.1 Attendees

Danny Bernstein

~~Andrew Curley~~

~~Bradley Daigle~~

Mike Durbin

Greg Jansen

Andrew Woods ★

### 4.9.2 Call-in Details

~~Dial in Number: (805) 399-1200~~

~~Participant Code: 900313#~~

*Let's try a new line*

Dial in Number: (805)309-2350

Participant Code: 6782943#

### 4.9.3 Agenda

1. Review initial [mock-up](#)
2. Define backend support data
3. Refine detailed requirements

### 4.9.4 Discussion

#### Balsamiq

- Be wary of potential bug with wiki plugin, apparently edits can be lost

#### Jeremy Boggs

- He will be at meeting on Oct. 30
- His exact skillset is unclear

## Mock-ups

1. Need separate runtime view of status
2. Need separate reports view
3. Need separate search view

## Runtime Reporting - Status

- Need to define exact runtime messages to display in "Activity Log"
- Are activity events captured by messaging? in a centralized log?
- Log will need to be normalized (json?) for display
- Make user-id visible in activity stream
- May need to have client tools supply metadata about a started 'ingest'
  - Decide on manifest?
  - Ingest should have a user-provided name
- Log messages will come from a variety of sources
  - IPS
  - DuraCloud events
- Hierarchy of ingest content
  1. Ingest event
  2. Packages
  3. Items
- Mike to spec out manifest
  - Tooling needs to create and send such a manifest
  - Manifest is sent before content
- Active Tasks
  - Clicking will potentially drill into detail
  - Progress bar is indication of % of items or packages of an ingest event
  - Details may include what files have landed, which are running
- Search on activity messages
  - Rely on time being the filter
- May not be necessary to capture packages/datastreams in Activity Stream

## Reports

- Discuss next time

## Next Meeting

- Thurs Oct 18th @10am ET



## 4.9.5 Actions

Description	Owner	Status
Updated Mock-ups	Danny	⚠️
Ingest Manifest Spec	Mike D.	some <a href="#">early specs</a> added
Detail Activity Log Messages	All	⚠️
Prepare to discuss Reporting & Discovery	All	⚠️

## 4.10 2012-10-18 - APT Admin Console Meeting

### 4.10.1 Attendees

Danny Bernstein

Mike Durbin

Greg Jansen

Andrew Woods ★

### 4.10.2 Call-in Details

Dial-in Number: (209) 647-1600

Participant Code: 700672#

### 4.10.3 Agenda

1. Continue review of [mock-up](#)
2. Define backend support data
3. Refine detailed requirements

## 4.10.4 Discussion

### Mockup - Dashboard

1. Storage graph
  - x-axis is time
  - y-axis is size
  - Remove "Ingests"
2. Elements on right of storage
  - Total of each visible space
3. Recent Activity
  - Include each time a checksum run happens (pass or fail)
    - Have a link to report
  - Ingest completed
  - Ingest cancelled
  - Future: delete
  - Future: running a report
  - Info icon
    - Downloads a text log
    - Failures need to be tracable, possibly an ! next to the info
4. Active tasks
  - Current progress on ingests
  - Info icon
    - Downloads a text log
5. Future: Pending tasks
6. Search
  - Will send user to Discovery page

### Mockup - Discovery

1. Is search on package or item?
  - Initial implementation will likely return package-level results
2. Need to be able to search APT metadata
  - Represented as facets
3. Facets
  - a. What is DPN bound?
  - b. What is world-readable?
    - Access controls: public/private/institution-only
  - c. Mimetype
  - d. Date ingested



## Mockup - Discovery Results

1. No need "More"
2. Need an item detail for a search item
  - Title
  - Ingest date
  - Files and checksums
    - packages/pids/datastreams
3. Checkbox per result to indicate download fixity report
4. Fixity details should be visually indicated at the package-level
  - Shows when the check was run
  - Need a simple backend query for getting checksum results
  - Potentially index fixity reports to support determining package fixity status

## Misc

- Noting that there are two usecases around packages
  1. It acts as the ingest unit
  2. It is also the scope of the search result

## 4.10.5 Actions

Description	Owner	Status
Clearly define "Content Discovery" questions (i.e. what is 'X')	<input type="checkbox"/> Andrew Curley	
Update required APT submission metadata	<input type="checkbox"/> Michael Durbin	

## 4.11 2012-10-30 - APT Admin Console Meeting

### 4.11.1 Attendees

Danny Bernstein

Mike Durbin

Greg Jansen

Andrew Woods ★

~~Jeremy Boggs~~

### 4.11.2 Call-in Details


Dial-in Number: (209) 647-1600

Participant Code: 700672#

### 4.11.3 Agenda

1. Bring Jeremy on-board
2. Continue review of [mock-up](#)

### 4.11.4 Previous Actions

Description	Owner	Status
Clearly define " <a href="#">Content Discovery</a> " questions (i.e. what is 'X')	<input type="checkbox"/> <a href="#">Andrew Curley</a>	

### 4.11.5 Discussion

#### General

1. Mike has made required APT metadata updates
2. Note to address to task of ensuring there are not PID collisions
3. Greg is organizing tech call with partners
  - Early Nov

## Mockup - Dashboard

1. Storage graph
  - Represents both spaces (staging and processed)
2. Active Tasks
  - May or may not include checksum events, depending on what data is available


## Mockup - Discovery

1. Search field / Indexing searches over
  - Rels-ext
  - Bit integrity reports
  - Possibly manifest
2. All packages will be shown if no query is input
3. Download bit integrity report
4. For errors, download error report

## Mockup - Package Detail

1. What is shown?
  - May need to show original content (DSpace objects)
2. Replace the left panel with object listing
3. Intent: prove that content has been ingested
  - Provide download of full object/datastream as one
  - Transform foxml for main panel (include MD5s)
4. Suggestion: on left panel show object "label" or something human readable
  - Include PID (handle or other)

## 4.11.6 Actions

Description	Owner	Status
Update <a href="#">Mockups</a>	<input type="checkbox"/> <a href="#">Danny Bernstein</a>	

## 4.12 2012-11-08 - APT Admin Console Meeting

### 4.12.1 Attendees

Danny Bernstein

Mike Durbin

Greg Jansen

Andrew Woods ★

Jeremy Boggs

Scott Turnbull

### 4.12.2 Call-in Details


Dial-in Number: (209) 647-1600

Participant Code: 700672#

### 4.12.3 Agenda

1. Bring Jeremy & Scott on-board
2. Continue review of [mock-up](#)

### 4.12.4 Previous Actions

Description	Owner	Status
Clearly define " <a href="#">Content Discovery</a> " questions (i.e. what is 'X')	<input type="checkbox"/> <a href="#">Andrew Curley</a>	

### 4.12.5 Discussion

#### Mock-ups

1. View is per institution

## Dashboard

1. Only one active task type: Ingest
2. Active Tasks
  - In-process activities
3. Recent Activity
  - Things that have completed
4. Rephrase 'packages' in Recent Activity to 'objects'

## Discovery

1. Solr in the backend
2. Will include search on PID
  - Result will show the package in which the PID is found
3. Warning sign links to error report
4. Need to revisit public/private access filter

## Discovery Detail

1. Package update needs to be further discussed
  - For example: may disallow replicated objects
  - For example: may delete all origin objects before second ingest
2. AuthN/Z of linked datastreams unclear

## UI Topics

1. Jeremy's initial focus
  - Usability
  - User interaction
2. Clarify terminology (may have different meanings across organizations)
  - objects
  - packages
  - datastreams
3. Ultimately Jeremy will create html and css
  - Jeremy has a slight preference for jquery / jquery-ui
  - CSS pre-processor frameworks can be helpful (compass & sas?)
4. Change "FOXML" link to represent downloading an archive of the whole object
5. Add Login view

## 4.12.6 Actions

Description	Owner	Status

## 4.13 2012-11-13 - APT Admin Console Meeting

### 4.13.1 Attendees

Danny Bernstein

Mike Durbin

Greg Jansen

Andrew Woods ★

Jeremy Boggs

Scott Turnbull


### 4.13.2 Call-in Details

Google Hangout

### 4.13.3 Agenda

1. Debrief technical webinar
2. [Mock-up](#) updates
3. Other technical updates

### 4.13.4 Previous Actions

Description	Owner	Status
Clearly define " <a href="#">Content Discovery</a> " questions (i.e. what is 'X')	<input checked="" type="checkbox"/> <a href="#">Andrew Curley</a>	



## 4.13.5 Discussion

### Webinar Debrief

1. Worked pretty well, and casual
  - UNC folks participated
  - Youtube resolution not good enough to read
2. 480 is the max free Youtube resolution
3. Consider using Twitter for comments
  - Also for links for slides, webpages
4. Capturing comments on wiki is a good history
5. First youtube comment should be a link to the wiki
6. Greg to move youtube video over to APTTrust to account
7. May be good to have regular technical updates
  - Invite participants to dream about the potentials
8. Useful to get some actual user feedback
9. Useful resource for user testing
  - <http://www.screencast-o-matic.com/>
  - Javascript
  - Shows user actions: mouse movement, clicks

### Mock-up

1. Mock-ups on the main wiki page will be the most current, child pages for older versions
2. Jeremy will be working on UI this Thursday
  - Has 2-3weeks of time to dedicate to this
3. Plan on having revisions every Monday

### Development Update


1. Mike is chunking content into DuraCloud
  2. Metadata datastream has been converted to RDF
  3. Mike to add others to github account
  4. Development work will be hosted in APT github account
  5. Danny to update the contacts chart to add github ids
- 
1. Before new year
  2. Mike and Andrew to hit deadlines

3. Jeremy to have mock-up shortly after T-Giving
  - Suggestion to push html in github
  - Google github-pages

## Other

1. This meeting will eventually evolve into the standing tech meeting

## 4.13.6 Actions

Description	Owner	Status
<a href="#">Greg Jansen</a> : Contact APTrust technical partners: <ul style="list-style-type: none"> <li>• Link to video</li> <li>• Why such little feedback?</li> </ul>	<a href="#">Greg Jansen</a>	

## 4.14 2012-11-13 - AP Trust Technical Update

The goal of the meeting is to share technical progress on phase one goals, answer your questions, and gather feedback.

[Live broadcast of the meeting will be available on this YouTube channel..](#)

### 4.14.1 Agenda

- [Technical milestones](#), Robin Ruggaber (AP Trust)
- Architecture review
  - Andrew Woods (DuraSpace) on AP Trust
  - Adam Soroka (UVA) on DPN
  - See the [architecture page](#).
- Ingest work flow
  - Mike Durbin (AP Trust)
  - See [ingest work flow](#) and [ingest client](#).
- Administrative interface
  - Danny Bernstein (DuraSpace) and Jeremy Boggs (UVA)
  - See [administrative interface](#) for use cases and scroll down for mock-ups.
- Closing, Greg Jansen (UNC)

## 4.14.2 Format

We are going to try hosting this via Google Hangout and YouTube. Google Hangout supports only ten participants at a time, but many more can interact via the live feed on YouTube. Here is how we expect things to go:

1. We will invite AP Trust presenters to take seats in the hangout.
2. When presenters are set up, we will begin the broadcast on YouTube.
  - a. YouTube live broadcast is tied to the hangout creator's channel. It will appear at approximately 1pm EST on [Greg Jansen's YouTube channel](#).
3. If there are still seats left, others with G+ accounts can join the Hangout.
4. Please submit comments and questions via YouTube or as comments on this page. We will respond to them in the broadcast.

This all worked in a limited test run, so hopefully it will provide a good format for our discussion. Let us know if you have suggestions.

## 4.15 2012-11-26 - APT Strategy Meeting

---

### 4.15.1 Attendees

Andrew Curley

Mike Durbin

Greg Jansen

Michele Kimpton ★

Greg Raschke

Robin Ruggaber

Donna Tolson

Tim Sigmon

Adam Soroka

Bradley Daigle

~~Andrew Woods~~

Scott Turnbull

### 4.15.2 Call-in Details

Dial-in Number: (805) 399-1200

Participant Code: 900313#

### 4.15.3 Agenda

1. End of Year Goals
2. Partner Institution Visits
  - a. Further questions for partners?
  - b. Take away items from visit to Duke, UNC & NCState
  - c. Note: Visit this week to Columbia and Syracuse.
3. Scott's first day, critical issues and updates.

### 4.15.4 Discussion

- APTTrust presentation at CNI to ask what should be covered at CNI presentation
- Responsibilities until end of December
  - Robin transferring all tech responsibilities for APTTrust to Scott. Getting accounts set up, finding positions
  - Scott getting picture of project. Will put together communications plan and partners list. Get more formal project plan together with end of year deliverables and April meeting. Who needs to be involved to get phase 2 underway. Schedule regular community communication plan.
  - Tim participating in weekly DPN phone calls and face to face at CNI
  - Adam working mostly with DPN. Design of architecture. Little bit of testing from APTTrust as node in DPN. Bring Scott up to date and how to make DPN fit with APTTrust.
  - Andrew C has been in holding pattern until revisit content issues. Participating in web console to advise as needed.
  - Donna- Martha and Donna finishing site visits to all partners before end of year and beginning of year. Working with Scott to figure out what he will be doing in outreach area, and kick up governance conversations and business models for continued participation in APTTrust.
  - Michele-working internally to manage the development team. Working with SDSC to get virtual compute spun up.
  - Mike-Ingest process routine, fedora staging to APTTrust fedora.
- Partner institution visits- Adam and Andrew sat down with Martha and Donna-will be visiting Columbia and Syracuse
- UNC,Duke, NCState visits have taken place
  - Everyone thinks what we have is great, but what we have developed thus far is not unique. Idea of building on deeper preservation services is very interesting. NCState very interested in preservation services, and about to set up fedora repository and would like not to stand one up on their own. Distributed preservation environment is real draw.
- Martha sent out brochure-on APTTrust
- Was clear preservation services is whole area of development very interested in. Want format migration, reporting format obsolescence. They want to control their own access repositories.
- One of the things they get from consortial effort is economies of scale, even if you are not interested in aggregating your content. These schools are serious about the TDR certification.

- Questions for Columbia and Syracuse:
  - Same as the initial questions in respect to phase 2
  - Not access services?
  - Validation of what other schools told you?
- What can Scott focus on that will get him in the game in a positive way, and get some blockages cleared? Other issues should be immediately aware of, what may be mid term issues?
- One of the midterm objectives is to assert the consortial model so their is partner contribution going on. Possibly code contributions, common problems identified.
- What are we asking partner institutions to be done on their end to move content to staging area. Need to still specify criteria for inclusion and minimal metadata to be provided before moving to APTrust.
- CNI presentations:
  - Survey results- what kinds of content will be deposited? Donna to provide clarification. This is an opportunity to set expectations,
  - Raising awareness
  - setting realistic expectations
  - laying groundwork what it means to be a member
  - being good candidates for this consortium-who is a good fit and who is not
  - community of practice-help create best practices
  - What APTRust offers
- APTrust is different certainly in services offered, and what is being created is a platform for consortial services- which can be preservation or access related.
- Projects to compare it to- Chronopolis, basic storage, Hathitrust, DuraCloud
- At what point do we open to new partners, and open up for general availability?
- In long term want to preserve our flexibility and ability to manage content with innovation
- Action items:
  - Donna- when do we open up for more membership? Updated survey data?

## 4.15.5 Actions

## 4.16 2012-11-27 - APT Admin Console Meeting

### 4.16.1 Attendees

Danny Bernstein

Mike Durbin

Greg Jansen

~~Andrew Woods~~

Jeremy Boggs ★

Scott Turnbull

~~Andrew Curley~~

### 4.16.2 Call-in Details

Google Hangout

### 4.16.3 Agenda

1. Boggs UI update ([AP Trust Administrative Interface](#) )

### 4.16.4 Previous Actions

Description	Owner	Status

### 4.16.5 Discussion

Update on Wire-frames for for UI.

- First pass for dashboard and discovery panel has been wire-framed.
- Consolidated notification list into single list for improved ease of reference and updates. Progress, failures and success is joined into one list.
- Institutional report improved to give a short summary of activity and counts for and individual institution as well as summary counts representing package facets.
- Add link to Institutional report indicating number of packages with Failed Heal Check.
- Links on the Institutional report summary accounts will bring them to a faceted browse for those items.

- Question about the need for what needs to be searched in the basic search box. PIDs for individual digital objects will be searchable but will return lists of the package those items are contained in.
- Administrative interface for APTTrust admin would be the same as each individual institutions. URI design for admin interface for individual institutions would be as follows "http://<domainname>:<port>/<institutioncode>/"
- Administrative APTTrust user would be able to access each individual institution through the same URI as above.
- The Package display block will include the user who initiated ingest and total size of files and date modified entry for packages that have been updates.
- Package Detail view will include views of the Fedora represented Datastreams in APTTrust, even if it originated in DSpace.
- Will be disallowing Digital Objects appearing in multiple packages as part of the overall ingest.
- Conversation about how to arrange project Git Repositories.
  - Prototypes should go in their own repository under the GitHub organization and will eventually be deleted when they're moved to the central Code Base.
  - Suggestion that we use a single Git repository for overall Java code-base for ease of updating and reference.

Big Item to discuss with technical group is what to do with audit trail and version of packages in the repository.

## 4.16.6 Actions

Description	Owner	Status
Finalize Wire-frames	Group	

## 4.17 2012-12-04 - APT Admin Console Meeting

### 4.17.1 Attendees

Danny Bernstein

Mike Durbin

~~Greg Jansen~~

Andrew Woods

Jeremy Boggs

Scott Turnbull ★

Andrew Curley

## 4.17.2 Call-in Details

Google Hangout 3-4 PM

## 4.17.3 Agenda

1. Any last UI updates (Jeremy - [AP Trust Administrative Interface](#) )
2. Final Wireframe Feedback (Group - Scott facilitate)
3. Template Creating Timeline Estimate (Jeremy)
4. Solr Installation, setup/config plan/estimate (Mike)
5. Other dependencies or requirements (Andrew/Danny)
6. Version Control repository (Group - Scott facilitate)
7. aptrust.duracloud.org server management (Andrew)
8. Greenhopper board. (Scott)

## 4.17.4 Previous Actions

Description	Owner	Status
UI Updates	Jeremy	

## 4.17.5 Discussion

**Finalizing Wireframe:** Tech group has no additional input and will open up to wider group for Thursday.

**Solr Instace:** Will start to setup this week and should be installed by end of week or beginning of next to work on sample data. Will be installed on aptrust.duracloud.org during the week. Documentation has been started about the Solr indexing fields and hooks will be added at a later date. SolrHome will be setup in the standard Tomcat install as a multi-instance with notes on installation process as needed.

**Other Dependencies or Requirements:** Rudiments of a working API and a definitive Git Repository. Authentication issues. Greg is contact for UNC and Scott will contact him to make sure no sensitive information is in there to keep us from using it as a development repository.

Can we create initial REST API for calls to the APTrust Ingest Code? Question about if this can be done by Solr API calls and store all needed information into Solr or if we need a separate API. As business logic needs crop up then questions of where to put these. Most of what we will want to call is detailed in Mikes code right now and he will work a bit beef up that detail to identify areas we need to do more with to get to work work.

Scott -> Add Danny and setup APTrust repo.



**Version Control:** Setup central APTTrust-Core GitHub repository to serve as the main version control repo for the project. Standard version control best practices are being documented [here](#) in the Wiki.

**aptrust.duracloud.org server:** Instance in being able to spin up server config so Andrew will look into setting up a scripted provisioning system. Will need to come up with some kind of server infrastructure needs and request additional support as needed (if we need more).

**Greenhopper Boards:** Don't feel there's a particular benefit for a Scrum or Kanban view for our current configuration. Regular JIRA for use with the project instead of a board.

Many people will be traveling to CNI next week so Danny and Mike will touch base if no formal meeting next week.

## 4.17.6 Actions

- ☐ [Jeremy Boggs](#) Create templates for Admin Interface and check-in to GitHub repo.
- ☐ [Michael Durbin](#) Install/Config Solr based on documented fields so far.
- ☒ [Scott Turnbull](#) Send out wireframe review meeting invitation to broader group.
- ☒ [Scott Turnbull](#) Make sure with UNC that no sensitive information is in the Repo for development.
- ☐ [Michael Durbin](#) Create simple web framework to include the endpoints needed to make non-Solr calls.
- ☐ [Danny Bernstein](#) Setup Initial APTTrust repository and add people with proper permission.

## 4.18 2013-01-08 - APT Admin Console Meeting

### 4.18.1 Attendees

Danny Bernstein

~~Mike Durbin~~

~~Greg Jansen~~

Andrew Woods

~~Jeremy Boggs~~

Scott Turnbull

~~Andrew Curley~~

### 4.18.2 Call-in Details

Google Hangout 3-4 PM

## 4.18.3 Agenda

1. Update On UI Development (Danny)
2. DSpace Content Workflow Initial Issues (Scott)

## 4.18.4 Previous Actions

Description	Owner	Status
Create templates for Admin Interface and check-in to GitHub repo.	Jeremy	In Process
Install/Config Solr based on documented fields so far.	Mike	In Process (stuck?)
Send out wireframe review meeting invitation to broader group.	Scott	Done
Make sure with UNC that no sensitive information is in the Repo for development.	Scott	Done
Create simple web framework to include the endpoints needed to make non-Solr calls.	Mike	In Process (stuck?)
Setup Initial APTTrust repository and add people with proper permission.	Danny	Done

## 4.18.5 Discussion

### Admin Interface Discussion

- Started integrating webapp to the Solr and processing scripts as well as the templates Jeremy used.
- Next week something preliminary to show in relation to the admin interface.
  - Will show webapp being run in part against the Solr instance.
  - We need to be flexible with staffing but if we need to pull in additional staffing if it becomes critical.
  - Aiming to have version 1 based on the features of the paper prototype by end Jan or early Feb.
- Solr index is not being used at this point in the prototype, we're using test data provided but nothing in terms of ingest yet.
- We're targeting Java/Spring skills in the new Software Engineer position to facilitate
- We have placeholders for Discovery and Package detail pages but need Jeremy to complete those before they can be implemented in the site.
  - Jeremy should have these done by next week.

### DSpace

- Thinking about using the following plugin for moving DSpace content to Duracloud - [ReplicationTaskSuite](#)
  - Generates AIP (archival information package) (oais) file and moves it into duracloud
  - Works with DSPACE 1.8 and above.
  - What are the inst partners running on?
    - NCSU 1.6 no plans to upgrade beyond 1.7
    - Duke 1.7.1 - want to jump to 3.1 directly when it comes out.
    - Johns Hopkins - 1.4.1 with no plans to upgrade.
  - How difficult to make it compatible with 1.7? Open question.
    - Perhaps we might be able to roll our own for AP Trust.
    - This replication tool won't create the ap trust metadata file anyway.
  - What about using [METS](#) format for the APTrust package metadata? (XML schema for dealing with structured sets of file and performing actions on them).
  - A replication tool is a probably a non-trivial task (according to Andrew's conversations with Tim) Donna was adamant about have backwards compatibility.
  - If all our partners are going to be on 1.7 and above, then AIP generation tool might be sufficient for our needs.
  - 1.6 supports command line access to the AIP generation tools.
  - Mike also mentioned that the DSpace objects will be much smaller than fedora objects. may or may not be true - it's possible to bundle many objects into a single AIP.
  - Large DSPACE items being ingested into APTrust: any ideas how to deal with it?
    - Why not use the chunker/stitcher?
    - Andrew agrees. nb: chunking must take place on the client side. Let's try to avoid having to deal with this scenario by setting a reasonable sub chunking file size and confirm with the partners that can agree to.

## 4.18.6 Actions

- ☐ Scott will get with Mike to break down specific tasks in JIRA so we can track and pick up work that may need to be done. (Do this at next Tuesday's meeting)
- ☒ Scott and Andrew will talk with Tim about the DSpace plugin and it's potential use for packaging and moving DSpace materials.

## 4.19 2013-01-15 APTrust Admin Console Meeting

### 4.19.1 Attendees

Danny Bernstein

Mike Durbin

Greg Jansen

Andrew Woods ★

Jeremy Boggs

Scott Turnbull

Andrew Curley

### 4.19.2 Call-in Details

[Google+ Hangout](#) 3-4 PM

### 4.19.3 Agenda

1. Update On UI Development (Danny)
2. Update on DSpace (Scott)
3. Update on Ingest Processing & Solr (Mike)
4. Next Tech Update? (Scott/Greg)
5. APT 1.0 technical timeline and task-breakout (Andrew)

### 4.19.4 Previous Actions

Description	Owner	Status
Create templates for Admin Interface and check-in to GitHub repo.	Jeremy	Done
Install/Config Solr based on documented fields so far.	Mike	
Create simple web framework to include the endpoints needed to make non-Solr calls.	Mike	
Talk with creator of Replication plugin for DSpace for applicability to APTrust	Scott/Andrew	Done
Update JIRA with tasks for current development or work.	Danny/Scott /Mike	

## 4.19.5 Discussion

### Webadmin

1. Partially being rendered by Java
2. No backend tie-in yet
3. Next: flush out security model (AuthN/Z)
  - Likely not need user-account-creation at this point
  - OAuth? Shib?
  - Eventually 12 partners, initially 6 or 7
  - Initially, web authN, store mapping of users to roles in-memory
  - Use Spring Security
  - Push Shib off for the moment

### DSpace

1. Version of institutional DSpace drives the mechanism of ingest (need 1.8)
  - JHU : 1.4
  - Duke: 1.7
  - NCSU: 1.6
2. Scott standing up UVa test DSpace 1.8
3. Would like to get Greg J. having conversations with partners around DSpace

### Ingest Processing Service & Solr

1. Solr is up with some test data
2. Some client code exposing Solr
3. Working on proposal for ingest manifest format
4. Curators' Workbench make be a useful tool for creating
  - package metadata
  - ingest metadata
5. Discussion around what "package" and/or "ingest" metadata should be
  - Manifests indicate what content is coming
  - What is the bottleneck? transfer time? or IPS?
6. Mike to put up a strawman package metadata format
7. Value of using METS, among other things allows inclusion of PREMIS events

## General

1. Email was received from member indicating a lack of communication
  - Push them towards aptrust website and wiki
2. Timeline and milestones should be established
3. Unclear what technical resources are available
  - Scott to come up with first cut by next week
  - Danny would like to get the bulk of the webadmin done by end of Jan.
  - As more work comes up, we need to determine has Danny availability
4. Ideally, we would have partner-based development as well
5. Start considering what new developers will (4-6 months out) will be doing
6. Use "components" in JIRA
  - fedora
  - dspace
  - admin-ui
  - admin-client
  - ips
7. Include non-dev tasks in JIRA

## 4.19.6 Actions

## 4.20 2013-01-22 - APTrust Admin Console Meeting

---

### 4.20.1 Attendees

Danny Bernstein

Mike Durbin ★

Greg Jansen

~~Andrew Woods~~

Scott Turnbull

### 4.20.2 Call-in Details

[Google+ Hangout 3-4 PM](#)

## 4.20.3 Agenda

1. Update on DSpace Planning (Scott/Greg)
2. Update on Ingest Processing & Solr (Mike)
3. Summary of Security Model Approach (Scott)
4. Update on Admin Interface (Danny)
5. Proposed Schedule and Format for Next Tech Update. (Scott/Greg)
6. Review of [Proposed Timeline](#) and Feedback. (Scott Facilitate)

## 4.20.4 Previous Actions

Description	Owner	Status
Install/Config Solr based on documented fields so far.	Mike	
Create simple web framework to include the endpoints needed to make non-Solr calls.	Mike	
Update JIRA with tasks for current development or work.	Danny/Scott /Mike	
Schedule Next Tech Update	Scott/Greg	Done
Setup DSpace Dev Instance	Scott	In Process
Outline Secirty Model for Admin Interface	Scott/Danny	Done

## 4.20.5 Discussion

### Notes

#### DSpace Update

Not much to report with DSpace. Greg and Scott worked to come up with a backwards compatible approach.

- Greg will schedule meeting with DSpace partners to discuss strategy and get buy in.
- Report to partners and get feedback if needed.
- Will report findings to this group after the meeting.

#### Ingest Processing and Indexing

- Had a productive conversation this morning around the concept of metadata packaging and how to create a manifest for processing files.
- Initial Fedora packaging was too dependent on a Fedora specific features and we wanted to abstract a method for everyone to submit metadata the same way.
- Ingest client will gather or infer information on packages to create manifest.
- The ingest client creates the manifest from metadata and sends it to the staging area.
- The ingest client parses the manifest package metadata then initiates a sync of the Digital Objects through either Cloudsync or ReplicationTaskSuite as appropriate.
- There are no implied changes to the admin interface by this method.

### **Security Model Approach**

- user logs via locally created and managed accounts on the admin client side.
- usernames, passwords and role are defined in a config file in the webapp.
- users may be super users or regular institutional users
  - institutional users can access content from their institution
  - super users can do anything

Danny requested clarification that the Solr Index would be firewalled to only allow access by the admin interface.

### **Admin Interface Update**

- changes will be checked in tonight or tomorrow
  - will include all mocked-up features
- after that, work will begin on security model

### **Schedule for next Tech Update**

- Scheduled for 3PM EDT February 14th (recorded google hangout)
  - Show a version of the Admin UI that can be loaded in a Browsers, but not using live data yet.
  - Update on DSpace strategy (meeting them at version 1.8 with plugin)
  - partner update from Donna will be included with the technical update.
- Donna suggested the content folks should be more engaged

### **Proposed Timeline**

- everyone considered the estimates to be reasonable and that we could hit a first of march alpha test
- we wish to test slightly more content for that test
- In mid February when we know what the ingest process will entail, we'll solicit partners to help prepare more content

In the various specific responses, there was a general consensus that the format of the meeting was working and that there was adequate communication, but that it was important to continue to evaluate the tools used. We may consider an alternative to Jira.



## 4.20.6 Actions

# 4.21 2013-01-29 APTrust Working Group Meeting

## 4.21.1 Attendees

- Danny Bernstein
- Andrew Curley
- Mike Durbin
- Greg Jansen ★
- Scott Turnbull
- Andrew Woods

## 4.21.2 Call-in Details

[Google+ Hangout 3-4pm](#) (Use link in Google Community Event if in doubt)

## 4.21.3 Agenda

1. Update on DSpace Partner Meeting (Greg)
2. Update on Ingest Processing & Solr (Mike)
3. Update on Admin Interface (Danny)
4. Tech Update Test On Air event invitation. (Scott)
5. Check-in on [Proposed Timeline](#) and Feedback. (Scott Facilitate)
6. Brief overview of [post-April timeline](#) Proposal (Scott)
7. Update on SE Interviews (Scott)

## 4.21.4 Previous Actions

Description	Owner	Status
Install/Config Solr based on documented fields so far.	Mike	
Create simple web framework to include the endpoints needed to make non-Solr calls.	Mike	

## 4.21.5 Discussion

1. Dspace Partner talk scheduled for Friday 10am
2. Ingest processing scripts
  - a. Should be ready for some kind of commit.
  - b. Format for serialization seems to be going along well.
  - c. Packaging XML modified to support entual multiple repos but this in scope for the demo.
  - d. Will work in a single manifest creation methodology with repo specific calls where needed.
3. Admin interface
  - a. Initial mockup demo was given with links working.
  - b. Items being displayed, pulling information from Solr.
  - c. Objects will display based on connection to Fedora currently.
  - d. Fedora objects to use for test have pid namespaces 'aptrust:x'
  - e. Security piece is the next to work out.
  - f. Danny will get with Jeremy to tighten up the look and feel for a better user experience.
  - g. Danny will start to work on the User Login Next.
4. Tech Update Test for On0Air. Greg will test with Scott
5. Demo timeline check-in
  - a. No further feedback.

## 4.21.6 Actions

- ☐ Danny will get with Jeremy to help tighten up some of the site look and feel.
- ☐ Danny will begin implementing the the simple User Login.
- ☒ Scott talk w/ Greg and Donna about moving Tech update so Danny can be there to Demo.
- ☒ Greg and Scott meet w/ DSpace partners on Friday to discuss version and planning.

## 4.22 2013-02-01 APTTrust DSpace Partner Discussion

---

### 4.22.1 Attendees

Scott Turnbull

Greg Jansen

### 4.22.2 Call-in Details

Google+ Hangout, link sent in invitation.

## 4.22.3 Agenda

1. Summary of DSpace content proposal.
2. Discussion of Version.

## 4.22.4 Minutes

Scott - 3 streams on incoming AP Trust imports

- fedora, dspace, file space (not in prototype)
- coming up with a DSpace plan for April prototype
- need a partner for April

Scott explained the basic AP Trust ingest work flow and packages

DSpace goal is to package exported AIPs

- various utilities exist
- the AIP import/export plugin only works for 1.8.1 (not reliable in 1.7)
- replication task suite plugin (cloud sync for DSpace)
  - syncs AIPs to DuraCloud space
  - would still need to find a way to move package metadata into the cloud

Can we find a partner for the prototype?

Is this method acceptable to partners as a longer term plan?

The remaining option of file-system preservation will remain once it is developed.

David Reynolds - John Hopkins

- big ETD project ongoing, publishing into a DSpace repo via Virio (new version of DSpace 3)
- no up soon enough for a prototype, but will work with end of year AP Trust DSpace work flow goal

Markus Wust - NCSU

- production is still 1.6
- test version of 1.8.2 on a VM with 3000 records, could be used in prototype time frame
  - PDFs articles and metadata
  - open to the public
- confirms he can supply AIPs to the cloud for prototype and some metadata

Jim Tuttle - Duke

- methodology sounds fine

- confirmed prior to 1.8.1 there was a bug in the bulk function of DSpace
- 1.7.1 in production
- all resource devoted to Hydra and Fedora as a migration from DSpace
- won't update DSpace again, just migrating
- if this plan doesn't come through they would then upgrade DSpace

Jim - Duke can help with the pre 1.8.1 DSpace testing when there is a call for that.

## 4.23 2013-02-04 APTrust Strategy meeting

---

### 4.23.1 Attendees

~~Andrew Curley~~

Mike Durbin

Greg Jansen

Michele Kimpton

Greg Raschke

Robin Ruggaber

Donna Tolson ★

Tim Sigmon

~~Adam Soroka~~

Bradley Daigle

~~Andrew Woods~~

Scott Turnbull

### 4.23.2 Call-in Details

Dial-in Number: (866) 842-5779

Participant Code: 6413798205

## 4.23.3 Agenda

### 1. Summary of partner visits (Donna)

[January2013 Update.pptx](#)

2. Roles for upcoming events: Feb. 15 partner call, April 29 full meeting (Donna, Scott, Gregs)
3. Date for in-person tech meeting in late Feb/early March (Donna)
4. Update on demo development plan (Scott) ( See PDF on the [Tech Milestone](#) page )
5. Update on DSpace Strategy (Scott)
6. Overall view of the post April [high level plan](#) (Scott)

#### Discussion

1. Partner visits almost complete - expect to talk to Michigan on 2/5/13. Good response from partner institutions, who are eager to get involved.
2. Upcoming events:
  - a. Feb. 15 call via Google Hangout: 3 parts - tech update (Scott/Greg J), partner visit summary (Donna), conversation about proposed services (Greg R). Goals for conversation are getting partners engaged and get going on setting priorities for future service development. Important to directly invite participation a few days ahead of time - Greg J will contact tech leads, Greg R will contact content leads.
  - b. April 29 partner meeting goals: show something "real" to partners, make concrete progress on choosing next steps (technical and service development), propose business plans for post-2013, position APTTrust so it begins to permeate partner institutions' culture and they can see what will be needed to align with APTTrust services. Donna will generate high-level outline for review and reaction. Important to have clear sense of who should come and benefit of attending, since it will take partner resources. Have basic meeting plan ready by February 15, share during Google Hangout.
3. Next in-person strategy meeting: first week of March out for DuraSpace. Donna will find a few dates that work with Martha and Scott and get something set up this week.
4. Plan basically on track for April demo. Internal demo target date is March 1.
  - a. Original milestones set in Sept. were unrealistic – see PDF link at top of Tech Milestone page for current deadlines.
  - b. Will update and upload new PDFs as plan is updated.
  - c. There are some items related to specific tasks still under negotiation that may appear on the plan including Server Administration and Setup and an expanded set of work for DPN.

5. Developing ingest tool for DSpace using ReplicationTaskSuite Plugin using the AIP export tool.
  - a. Requires DSpace version 1.8.1, which no partners have yet.
  - b. NCSU will be setting up a DSpace 1.8.1+ and some test content for using in development and to use in the April Demo.
  - c. Greg & Scott have talked to 3 DSpace partners about issue, and all three are ok with this approach for testing – they will either update soon or move to Fedora.
  - d. See the [meeting notes](#) for more information.
6. Overall view is not firm yet – just a proposal to react to.
  - a. Basic plan is to focus on demo for April, then mature processes for production, then focus on service development.
  - b. Requires certain staffing levels outlined in resource tab.
  - c. Not so much reworking code as expanding on what has been developed in start-up phase.

## 4.24 2013-02-05 APTrust Working Group Meeting

---

### 4.24.1 Attendees

- Danny Bernstein
- ~~Andrew Curley~~
- Mike Durbin
- Greg Jansen
- Scott Turnbull ★
- Andrew Woods

### 4.24.2 Call-in Details

[Google+ Hangout](#) 3-4pm EST

### 4.24.3 Agenda

1. Update on DSpace Partner Meeting (Scott)
2. Update on Ingest Processing & Solr (Mike)
3. Update on Admin Interface (Danny)
4. Tech Update Test On Air event invitation. (Scott)
5. Check-in on [Timeline](#). (Scott Facilitate)
6. Brief Update on SE Interviews (Scott)

## 4.24.4 Previous Actions

Description	Owner	Status
Consult w/ Jeremy to improve site look at feel.	Danny	In process
User Login on Site	Danny	In process
New Date/Time for Technical Update	Scott/Greg	Done
Get Agreement and Development Partner from DSpace Partners	Scott/Greg	Done

## 4.24.5 Discussion

1. Update on DSpace Partner Meeting (Scott)
  - a. DSpace partners fine with strategy of using ReplicationTaskSuite plugin on version 1.8.1+
  - b. Markus Wurst with NCSU is setting up a test instance there with ~35k test records for us to use in the demo.
  - c. Plugin should be setup and that environment ready for when we want to move toward packaging and processing Dspace content.
2. Update on Ingest Processing & Solr (Mike)
  - a. Initial ingest scripts checked into github.
  - b. Next step will be to create code to process fedora packages.
  - c. Strategy for ingesting Fedora objects is to bring them from staging using Cloud Sync and rebuilding the original Fedora objects in our repository to give access to datastreams. Will also allow us to use native checksums on those objects.
  - d. Seem to be on track with current schedule as far as everyone can tell.
3. Update on Admin Interface (Danny)
  - a. Another check-in pending for the security piece of the application.
  - b. Haven't spoken to Jeremy yet for UI tweaks and such.
  - c. Using config file authentication method where user accounts are given roles there.
  - d. Users will be added to the XML of the spring config.
  - e. Institution IDs need to be supplied,
4. Tech Update Test On Air event invitation. (Scott)
  - a. Mike asked if we can get questions
5. Check-in on [Timeline](#). (Scott Facilitate)
  - a. Timeline mostly seems on track. No other input.
6. Brief Update on SE Interviews (Scott)
  - a. General update given, particulars not appropriate for this format.

## 4.24.6 Actions

- ☒ Scott - Get a list of institutions codes to be used throughout the application.
- ☒ Scott - Consult with Danny on build process for the client app.
- ☐ Scott - Distribute Tech talk schedule to tech team.
- ☐ Danny - implement login in the admin application.
- ☐ Danny - Get w/ Jeremy Boggs for UI tweaks and improvements.

## 4.25 2013-02-14 APTrust Working Group Meeting

---

### 4.25.1 Attendees

- ~~Danny Bernstein~~
- Andrew Curley
- ★
- Mike Durbin
- Greg Jansen
- Scott Turnbull
- Andrew Woods
- Dan Davis

### 4.25.2 Call-in Details

Google+ Hangout [link is in the event](#).

### 4.25.3 Agenda

1. Update on development & any critical issues. (Mike)
2. Overview of Partner Update (Scott)
3. DPN Update (Scott)

### 4.25.4 Previous Actions

- ☒ Scott - Get a list of institutions codes to be used throughout the application.
- ☒ Scott - Consult with Danny on build process for the client app.



- ☒ Scott - Distribute Tech talk schedule to tech team.
- ☐ Danny - implement login in the admin application.
- ☐ Danny - Get w/ Jeremy Boggs for UI tweaks and improvements.

## 4.25.5 Discussion

- Dan Davis has joined this conversation for the first time. Long-time committer for Fedora. Worked with Duraspace and Fedora common
- Scott wants to test Replication TaskSuite with DSpace and NCSU. Perhaps should rope Tim in for this case study. Marcus is our main point of contact at NCSU though.

## Development Update

- Mike is continuing work on the Ingest Processing Services. Worked on client code that hits DuraCloud. Plans to get it done by tomorrow (2/15).
- Scott worried about hard-coded namespaces. They are RDF assertions that DPN might use to query. Does the packaging service need namespaces?
- We need to continue fleshing out the preservation package model.
  - Greg recommends using a namespace of a domain you control. aptrust.org?
  - Scott wants to support versioning and therefore wants to worry about this later.
  - Mike agrees to parameterize this.
  - Should we have a web cache for allowing other to validate XML against our schema.

## Partner Update

- Scott will update the partners. Highlight work that DuraSpace did. Work that Mike did
- Broach the need to create Policy committee.
- Danny will be talking about his work as well.
- Donna and Greg Jansen will also be participating.
- Going to do an on-air Hangout. Never done this before.
- Some concerns about how comments will be made by the public.
- General discussion of logistics.

## DPN Update

- Met in Florida with most of the DPN technical and management team. Impressed with enthusiasm.
- We are coming up with a set of policies for messaging. Will have to move conversation to packaging policies. Will be based on BagIt.
- Encryption? Versioning? How do we handle these? We need a policy.
- We need to have a semi-functioning application by the fall.

- Andrew W says that they will have more than a beta but less than production application and system by Fall 2013.
- Right now there are three access policies but there is no where in the APTrust model that we can handle real rights policies.
- We will probably have a blanket institutional policy rather than object by object.
- Scott thinks we need a policy store on the APTrust end. Have a validated step to confirm that object can go or not.
- Scott is working on the message queues with Andrew W. Will not get to work on it until next week though.
- Currently five nodes and each are part of DPN because they are trusted archives. Mostly different implementations. Some touchpoints.
- Scott is writing APTrust's DPN code in Python.

## Miscellaneous

- Want us to start thinking about the post-April future. BagIt in CloudSync? DuraCloud? What is the architecture ongoing?

## 4.25.6 Actions

## 4.26 2013-02-19 APTrust Working Group Meeting ---

### 4.26.1 Attendees

- Danny Bernstein  
Andrew Curley
- Dan Davis
- Mike Durbin  

- Greg Jansen
- Scott Turnbull
- Andrew Woods
- Markus Wurst

### 4.26.2 Call-in Details

Google+ Hangout see [Event Page](#) for hangout link.

- Check in on Web UI login and tweeks (Danny)

- Fedora Ingest and Processing update (Mike)
- Check in on plan schedule. (Scott)
- Plan for ReplicationTaskSuite test (Scott/Markus/Dan)
- Schedule Post April Architecture feedback session w/ Group (Scott)

### 4.26.3 Previous Actions

- ☐ Danny - implement login in the admin application.
- ☐ Danny - Get w/ Jeremy Boggs for UI tweaks and improvements.
- ☒ Group - Partner Update.

### 4.26.4 Discussion

- Dan Davis and Markus Wurst were present.
- Markus would work on DSpace and replication task suite

## Meeting Notes

- Web UI Update
  - Danny spoke with Jeremy who is quite busy with other projects right now
  - Danny made some tweaks before the demo
  - It may be premature to polish things up right now, especially considering his time constraints
  - Curious to hear feedback from the demonstration last Friday
    - Scott hadn't really heard any
    - Nor had Greg J
  - Scott asked about the login functionality
    - Danny indicated that it was there: partner logins and a root login, would verify that it has been committed
  - Aside from cosmetic pieces security is done.
- Fedora Ingest and Processing
  - In process of updating this to get it to work.
  - Required Solr changes
  - Planning on checking final Fedora Ingest scripts in next few days.
  - At that point Mike will deploy the service so that if files go there it will go into Solr index.
  - Mike will notify Danny when he's done with processing scripts so Danny can change the data reference points in the webapp
- Check in on Plan Schedule
  - Mike indicates that he'll be able to finish his updates this week and other milestones won't need to be pushed back
  - Scot will update the plan to reflect the actual completion dates

- Update on Replication Task Suite and DSpace
  - An NCSU staging and production space will be set up
  - Scott will get together with Markus to test pushing all ETDs to staging area (small files, but about 3GB of total data)
  - Dan doesn't anticipate difficulty setting up spaces
  - Mike will send a sample XML manifest to Markus
- Post-April Schedule Architecture Feedback
  - want to get feedback quickly to
    - determine tech stack
    - create appropriate job descriptions
  - may need to find contractors, temporary help between now and when the new staff is hired
  - Thursday, 21st at 4:00PM EDT
  - Big questions have to do with how we package and manage stuff
  - Other answers depend on the policy group (not yet formed)
  - There is a need for a "scenarios group"

## 4.26.5 Actions

- ☐ Mike should send Markus a sample manifest for DSpace content
- ☐ Those who are interested and haven't already discussed the post-april plan should meet on Thursday
- ☐ Mike will work with Danny to work through integrating the client with the admin interface

## 4.27 2013-02-25 APTrust Strategy Meeting

### 4.27.1 Attendees

Danny Bernstein

~~Andrew Curley~~ (tried to call in)

Bradley Daigle

Dan Davis

~~Mike Durbin~~ (tried to call in)

~~Greg Jansen~~ (dropped for most of call)

Michele Kimpton

~~Greg Raschke~~

~~Robin Ruggaber~~

Tim Sigmon

~~Adam Soroka~~

Donna Tolson ★

Scott Turnbull

Andrew Woods

## 4.27.2 Call-in Details

Dial-in Number: (805) 399-1200

Participant Code: 900313#

## 4.27.3 Discussion

### Feb. 15 Hangout Debrief (Scott, Greg J, Greg R)

- Seemed to go well, despite a few technical glitches
- No solid info on who participated or viewed afterward, but we know several people have
- Stay with Hangout format for now so partners become used to the technology, but perhaps revisit with group on April 29

### In-person tech team summit prep - March 11-12

- Logistics (Donna)
  - Room TBD, hopefully Scholars' Lab classroom
  - Begin with lunch at noon on Monday
  - Working session until 5pm
  - Groups for dinner – yes, interest in this.
  - Working session begins at 9am
  - Lunch provided
  - Finish up by 2pm
- Goals (all) (Proposed):
  - Clarity on roles & expectations: update on overall project trajectory, UVa & DuraSpace contributions, core APTTrust staff, contributors from partner institutions
  - Preparation for April 29: review work needed, responsibilities and timelines, outline technical presentation
  - Setting stage for future:
    - Completion of Phase I - needs clarity:
      - business model defines this as through Dec. 2013
      - tech model considered Phase I period leading up to demo
      - also need to think about when we cross line into operationalization
    - Preparation for Phase II
      - what will business model look like beyond Phase I?
      - Could APTTrust serve as pass-through service for those interested only in DPN?

- Agenda
  - Scott will draft initial ideas during tech team call on Tuesday 2/26, Donna will refine after 3/7 advisory group meeting
  - Include time for breakout groups (tech team, planning team, etc.)

## Questions for Advisory Group? (March 7 meeting)

Formation of a Policy group.

- Mix of Technical, Archival and Governance skills.

Use Case Scenario Teams

- Overall use case examples.
- Cut down list based on actual scope of project
- Target actual cases we want to address based on impact and opportunity.

Membership

## 4.28 2013-02-26 APTrust Working Group Meeting

---

### 4.28.1 Attendees

- Danny Bernstein
- Andrew Curley
- Dan Davis
- Mike Durbin
- Greg Jansen ★
- Scott Turnbull
- Andrew Woods
- Markus Wurst

### 4.28.2 Call-in Details

Google+ Hangout see [Event Page](#) for hangout link.

### 4.28.3 Agenda

- Check in on Web UI login and tweeks (Danny)

- Fedora Ingest and Processing update (Mike)
- Feedback on ReplicationTaskSuite (Markus)
- OR Proposal Update (Scott)
- Tech Team Summit Call for Agenda Items (Scott lead)

## 4.28.4 Previous Actions

- ☒ Danny - implement login in the admin application.
- ☒ Danny - Get w/ Jeremy Boggs for UI tweaks and improvements.
- ☒ Mike should send Markus a sample manifest for DSpace content
- ☒ Those who are interested and haven't already discussed the post-april plan should meet on Thursday
- ☒ Mike will work with Danny to work through integrating the client with the admin interface

## 4.28.5 Discussion

Mike - fedora ingest updates

- in a couple days he will complete the ingest and processing code

Andrew recommends we use direct login for authentication, instead of RSA certificate authA

Scott

- 15th for first run at the demo

NCSU DSpace Replication Task Suite

- will try installing the manifest building script
- was able to sync several items to the NCSU cloudspace
- actual processing was just a few seconds, just had to pick the items from a drop down
- Andrew - did existing docs cover the bases?
- Markus
  - no larger issues, most of the difficulty was in setting up the DSpace in the first place.
  - will try to add some comments to the configuration file for clarity
- Scott - debrief by Mid-march to create a documented DSpace workflow

Scott is moving this weekend

Proposal for OR on AP Trust architecture

- highlight use of community-based tools
- unifying view/use of tools - DuraCloud, Sync Tech Stack, Hydra stack
- branch out in terms of sets of services and how specialist manage preservation content

- draft will be out later in the week for review and suggestions
- due on Monday EOB

Tech Team meeting

- agenda ideas, see "Tech Team Summit March 11-12 2013" google doc
- Scott - enough ideas collected to fill the schedule

## **4.28.6 Actions**

## **4.29 2013-03-05 APTrust Tech Summit Meeting**

---

### **4.29.1 APTrust Tech Summit Notes**

Notes were taken in raw form along with the agenda.

### **4.29.2 Monday**

12pm - Summary of Goals: Introductions & Reiterate Project Overview - Donna (~1 hr)

- Light Lunch served.
- Introductions: Name, skills, involvement - Group (~ 30 min)



- Martha Sites (~20 min)
  - Needs to clearly hear the blockers for the project so we can identify a strategy to move forward.
  - Director position posted and reviewing applications.
  - Will be bringing on 2 positions for APTrust/DPN to help move the project forward.
  - Several partners, one in particular, is very interested in directly committing resources to help develop some features of APTrust.
  - Need to stay focused on enriched features to make a compelling case for value in the effort for our partners.
  - Need to keep the discussion bi-directional so everyone has consistent information and we all understand what is being reported out to partners.
  - We're beginning a criteria to determine how we'll bring in new partners or institutions into APTrust. We want to grow the community but it's not immediately clear what that exact number is to balance growth with a manageable community.
  - Expansion and sustainability will be heavily dependent on participation by deans and presidents.
  - We'll need effective involvement from the tech team on helping to shape the strategies and policies involved in the project both from a practical point of view but also from a sustainable workflow and best practices point of view.
  - The aim of the repository keeps content managed entirely by the original institution but to leverage what we can from aggregating content centrally. The security requirement is that content management is isolated entirely to the originating institution.
- Michele Kimpton (~10 min)
  - Interest in helping to keep APTrust community focused and to collaboratively develop strategies for evolving the community of tools together.
  - Hearing the overlap of needs between stake-holders in the community helps uncover broader needs across the community.
  - DS stewards Fedora in the community and helps connect the community and moves it along in a collaborative fashion.
  - Might be able to help us move needs development requests into the Fedora or DSpace code.
  - Duracloud service and the greater use of this by the community as a whole helps fund and improve the the toolset for everyone.
  - DS is moving to help enhance the open source offerings to diversify the offerings they make to the community as a whole and enhance their ability to connect participants into the greater community of practice.

#### 1pm - April Demo Overview & Planning - Scott (~1 hr)

- April 29th in Chapel Hill.
- Presenting to Deans Tech Leads and Content Leads from each of the Institutions. We anticipate most partners will send some representatives
- Agenda not set yet and it will be a full day. Likely single format presentations and break out sessions.
- Want to open to remote viewing of most sessions but only broadcast to partners and not to the public as a whole.
- ACTION ITEM: Donna will get with venue host (Greg J.) to find out if broadcast is practical.

- Some combination of Functional Demo, Technical Deep Dive and Policy Deep Dive.
- Will have a technical demo:
- High level items should be focused either in the morning or afternoon so Deans can focus their time best. We think they are most interested in:
  - Very high level overview.
  - General Policy.
  - How to influence Governance.
  - Cost Models of high interest to them.
  - Will want to have some level of influence on the content aspects of the repository. Should it be focused on collection dev, format?
  - ACTION ITEM: Greg R. should present his plan for how to develop a plan and prioritize service development in APTrust.
  - Will want to provide enough outline to provide the boundaries of a reasonable discussion, particularly on content management.
  - Advisory group will discussion initial ideas for how to expand membership.
- Question of what expanding membership means?
  - Criteria for expanding membership will be determined by steering group.
  - Need to understand what it means to become a member? Help with development? What are the impacts on production.
  - There are a number of barriers that very unique per partners so we need some sort of expanded definition of when we can bring services.
- What does this group think needs to be done here.
  - Should try to aim for mixed composition groups in the breakouts to foster discussion.
  - Deans might need to opt out after morning.
  - Post April plans are and establish a reliable method to communicate progress.
  - Cross-checks on our Policy and Scenarios as they are developed so far.
  - Separate Group or breakout session for security scenarios:
    - Data Protection Agreement essentially defines requirements for storing content off-site. Get some way to centralize this agreement.
  - Stress in our discussions the streamlining of governance and a single place to have that discussion.
  - Go over the DPN side of APTrust, the implications and things they may need to do based on our current understanding. Deletion specifically presents problems.
  - Methods for managing preservation decisions and evaluating it's place in the APTrust/DPN echo system.
  - Tech group needs to feed up costs for service, IO, servers, and staffing.
  - Cost model should be very simple and something we can modify easily.
    - Tech group should feed up known costs.
  - Martha, Greg and Donna will discuss more about cost modeling with the group.
  - Perhaps conversation about Service Level Agreements.

- Technical Demo and how that will go:
  - Hour Demo.
  - Restate Phase I goals.
  - Overall architecture and workflow.
  - Initiate a transfer from repository into the Admin interface from both DSpace and Fedora.
  - Fallback of process via video in case server balks.
  - Restate value of centralize process and structure if not already included.
  - Lessons learned from developing proof-of-concept.
  - Demonstrate the lense to the preserved packages after they are ingested.
  - What happens on a failure.
  - How to view original content as part of a preservation package.
  - Summarize how this position us to grow services more and develop throughout the year.
  - Verbally address the idea of how non-fedora/dspace content will be handled.
  - Some view of the underlying tools
  - Suggested workflow and how the preps them on technical side and what the expectations on the partners is.
  - Demo -> architecture -> Demo.

2pm - Brief DPN Overview & Implications for APTrust - Scott (~ 30 min)

- Scott presented during this and couldn't take notes himself.
- Gave a basic overview of DPN.

2:30pm - Break

3pm - Suggest guidelines and priorities for Policy and Scenarios Groups - Greg (~1 hr)

- Discussion centered around checksheet provided by Greg j [https://docs.google.com/spreadsheet/ccc?key=0An\\_Ujm5dqTQRdFNoc21OeGpPck5fdUJ4aGYyWmFtT0E&usp=sharing](https://docs.google.com/spreadsheet/ccc?key=0An_Ujm5dqTQRdFNoc21OeGpPck5fdUJ4aGYyWmFtT0E&usp=sharing)
- General set of recommendations on what we think Greg R's group could focus on early to help with things.
- Notes taken directly in the Google Spreadsheet.

4pm - Defining Roles and Responsibilities - Donna (~30 min)

- There are many different contexts within the project to define a role and it makes it difficult to reach specific definitions.
- Breaking things down more by a goals.
  - Develop Preservation Repository.
  - Establish high functioning collaborative community.
  - "Raise the flag" for digital preservation advocacy.
  - Maintain operations and emergency response.

- Qualities
  - Sustainability
  - Collaboration
  - Leadership
- This meeting mostly focused on Building/Development and on Operations Management.

4:30pm - Assess Phase I Goals - Greg (~1 hr)

- Package Creator (CRU) via web interface and api.
- DPN package creation and management.
- Return Original Objects to institutions.
- Deliver Supplemental information about objects.
- Digital Object Auditing and Reports.
- Aggregate Content Reporting.
- Ingestion workflows and support for Fedora, DSpace and Flat File Digital Objects.
- Add, Update or Delete Digital Objects to a Package.
- Audit Services on Packages and associated Digital Objects.
- Operations Management & Operational Plan.
  - Service Level Agreements
    - Access
    - Redundancy
    - Content Management Agreement.
  - Service Monitoring.
  - Emergency Response.
  - Application Administration.
  - Systems Administration/Update
  - Tech Support
  - Security Auditing and Management.
  - Quality Assurance, (error log checking etc.)
  - Account Management

\* We'll be organizing dinner on the Charlottesville Mall this evening for anyone who would like to go.

## Tuesday

9am - Packaging Workflows - Scott (~1 hr)

- Method of designating content set for a cloudsync is TBD and may depend on partners, some kind of SPARQL, perhaps marked objects, perhaps other

- ways to lock staged files during ingest transactions:
  - have CloudSync set to not overwrite files in DuraCloud, to prevent changes during an ingest. (This depends upon well behaved users.)
  - lock the entire space
  - use DuraCloud attributes to track state
  - overwrite is not be a problem for Fedora objects, since data streams always have a specific version, so edits will not collide (need to test this)
  - another option is just to fail the ingest processing if items change during the transaction
- Mike: Lots of ingest/update decisions need to be made with more input.
  - example of moving a PID from package A to package B

#### 10am - BagIt & Versioning Initial Planning - Scott (~1 hr)

- Versioning Assumption: We keep versions of our own aptrust metadata, but we are agnostic towards the partner digital object, i.e. we keep the current copy and all versioned data within it.
- Content models: Ingest transactions + package objects + bags (as required by use case) (institutional repository objects)
- Bags are agreed upon for, 1 for each original repository object
- Package objects reference the bags, either as external datastreams or in RELS-EXT.. (External data stream gives us some metadata slots for checksum, etc.. and API retrieval.)
- Bags will be referenced as external datastreams within package objects (see above) and can be exported.
- Bag delivery is near term export function. Respond to further content/scenarios group needs as they emerge.

#### 11:00am - Initial Plan and Schedule Post April - Scott (~1 hr)

- Whiteboard picture of epic user stories goes here  
See picture in original google doc: [https://docs.google.com/document/d/1nNU5hURVoRdQW\\_Qa5gXVny2hgX03t0MpGIVCIXIJvQc/edit?usp=sharing](https://docs.google.com/document/d/1nNU5hURVoRdQW_Qa5gXVny2hgX03t0MpGIVCIXIJvQc/edit?usp=sharing)

#### 12pm - Initial Systems Architecture and Hosting - Scott (~30 min)

- 2 or 3 virtual servers
- DC as a separate machine.
- JMS turned off in Fedora, not needed our architecture.
- 2 things to achieve.
  - Service runs in a way that gives a good picture of basic performance without load.
  - Ability to monitor and gather data to make prediction of server needs for production.

- 2 Instances of at least large reserved (1 more small for DPN):
  - S3 for Duracloud
  - Elastic storage for individual servers and temp space.
  - Duracloud Only Instance for one server; Everything else on 2nd server.
  - Additional small instance server for DPN.
  - Questions about Development access to services and servers. This will be handled on a case by case basis between devs and Dan.
  - Setup via individual developer accounts as needed within tools.

12:30pm - Lunch Served (~30 min - continues through next session)

1pm - Wrap Up & Next Steps - Scott (~1 hr)

- Greg R's Policy and Scenarios group will get our input on policy priorities - group call with Scott, Donna, Greg J.
- Scott will formulate the project plan based on deliverables list
  - Some tasks will fall out for policy development
- Post 2 positions for the core APTrust team.
- Identify other resources available for parts of the development plan.
- Post developer role needs for partner contribution

## 4.30 2013-03-05 APTrust Working Group Meeting

### 4.30.1 Attendees

- Danny Bernstein
- Andrew Curley
- Dan Davis
- Mike Durbin
- Greg Jansen
- Scott Turnbull ★
- Andrew Woods
- Markus Wurst

### 4.30.2 Call-in Details

Google+ Hangout. See [Event Page](#) for Details

### 4.30.3 Agenda

1. Fedora Ingest and Processing update (Mike)
2. Feedback on ReplicationTaskSuite (Markus/Scott)
3. [OR Proposal](#) Update (Scott)
4. Update on Tech Team Summit (and Steering Group Meeting?) (Scott) [[Notes Link](#)]
5. DPN Update (Scott)
6. More?...

### 4.30.4 Previous Actions

- ☒ Scott - Submit OR Panel Proposal.
- ☒ Scott - Prepare Draft Tech Team Summit agenda.
- ☐ Mike - Complete Fedora Ingest code and connect to admin interface.
- ☐ Dan - Work with Mike to clear up login issue to staging server.
- ☒ Markus - Run ReplicationTaskSuite test set of DSpace content.

### 4.30.5 Discussion

1. Fedora Ingest and Processing update (Mike)
  - a. About 95% done with code for ingest processing services.
  - b. Currently concentration on integration testing against real data.
  - c. Overall complexity is higher than we anticipated mostly due to trying to figure out if an object is "there" or not.
  - d. The take forward message is mostly that there is a bit more complexity involved in validating the objects as moved from Fedora.
  - e. We may need to modify Cloudsync to include a mechanism to indicate when a Digital Object is finished being copied. Perhaps we need to add something to an end manifest.
  - f. Goal is to be able to see something for the tech summit by next week.
  - g. Mike will document issues related to IDs to make them be unique across all object IDs.
  - h. Danny is close to being able to tie into the back end and wanted to know from Mike if he can hookup his code so far.
    - i. Solr right now needs to be reconfigured before Danny can integrate the index effectively.
    - ii. Danny and Mike will work out connecting the back-end.

2. Feedback on ReplicationTaskSuite (Markus/Scott)
  - a. Markus installed a test instance of
  - b. 30 items per minute and these were small items in the range of several K.
  - c. Wants to run a test just produce AIPs normally through DSpace and use SyncTool to move them to the them to Duracoud Instance.
  - d. After running comparison we can talk to plug-in creator to get some advice on speeding the process up.
3. [OR Proposal](#) Update (Scott)
  - a. Proposal Submitted on Monday for a Panel.
  - b. In process of contacting potential Panel members (Scott, Greg J, Adam S, Michele or Andrew, more? different?)
  - c. Will finalize composition of panel as we approach OR.
  - d. Will use OR to try to pull together a meeting/discussion for potential partner interest in development.
4. Update on Tech Team Summit (and Steering Group Meeting?) (Scott) [[Notes Link](#)]
  - a. Schedule has been updated and in need of review.
  - b. No specific feedback on schedule.
  - c. Will get feedback from planners who may make changes.
5. DPN Update (Scott)
  - a. Focus is on MQ implementation right now with a lot of coding by Scott.
  - b. Other planning underway for Versioning and BagIt Strategy that may impact APTrust.

## 4.30.6 Actions

- ☒ Mike will finish up the Ingest scripts and hopefully and connect to the admin interface this week.
- ☐ Danny will get access to Duracloud from Dan Davis.
- ☐ Markus will run a test using SyncTool against AIPs to compare workflows.
- ☒ Mike/Danny will work together to get initial data connected to the Admin interface on Thursday.

## 4.31 2013-03-19 APTrust Working Group Meeting

### 4.31.1 Attendees

- Danny Bernstein
- Andrew Curley
- Dan Davis
- Mike Durbin
- Greg Jansen
- Scott Turnbull



- Andrew Woods ★
- Markus Wust

## 4.31.2 Call-in Details

Google+ Hangout link is on the [event](#) invitation.

## 4.31.3 Agenda

1. Ingest Client. Documentation and Configuration needs (Mike)
2. Review important policy needs for policy group. (Greg)
3. Schedule Initial Demo of Admin Interface. (Scott)
4. Starting up DPN related issues. (Scott)
5. Curator's Workbench
6. More?

## 4.31.4 Previous Actions

- ☒ Mike will finish up the Ingest scripts and connect to the admin interface.
- ☐ Danny will get access to Duracloud from Dan Davis.
- ☐ Markus will run a test using SyncTool against AIPs to compare workflows.
- ☐ Dan - Work with Mike to clear up login issue to uva.duracloud.org server.

## 4.31.5 Discussion

### Ingest Client

1. Tying admin console to backend works
2. Mike to document the pieces this week
3. Mike to deploy this week: [aptrust.duracloud.org](http://aptrust.duracloud.org)

### Scheduling Initial Demo of Admin Interface

1. April 4th @3pm
2. Pieces to finish by demo
  - a. IPS needs to run as a service
  - b. Need to add code to help with content ingest
  - c. May be some code on Danny's side to complete

3. Add initial user stories to Pivotal Tracker
4. This is just an internal demo

## Review important policy needs

1. There is some confusion that DPN is the preservation repository for APTTrust
2. Policy questions should not be limited to TDR
3. Link google doc to wiki
4. Reviewing and recording issues in Policies Needed by Technical Team document

## DPN Related Issues

1. Need to form a group to discuss DPN issues
2. First need to understand DPN and APTTrust differences in mission
3. What constitutes in APTTrust, a DPN item
4. Need to help partners understand when to push content to DPN or not

## Curator's Workbench

1. Bags should likely be at digital object level
2. Soon after April demo get into requirements and use cases
3. Potentially rope in other people at OR13
  - Tie in DuraCloud integration at Curator's Workbench workshop

## Fedora4

1. Interest in targeting Fedora4 and Hydra

## 4.31.6 Actions

- ☐ [Michael Durbin](#) to deploy IPS to [aptrust.duracloud.org](https://aptrust.duracloud.org)
- ☐ [Michael Durbin](#) to document deployment/configuration of pieces running on [aptrust.duracloud.org](https://aptrust.duracloud.org)

## 4.32 2013-03-21 APTrust Strategy Meeting

---

### 4.32.1 Attendees

~~Danny Bernstein~~

~~Andrew Curley~~

Bradley Daigle

Dan Davis

Mike Durbin

~~Greg Jansen~~

Michele Kimpton

~~Greg Raschke~~

~~Robin Ruggaber~~

~~Tim Sigmon~~

~~Adam Soroka~~

Donna Tolson

Scott Turnbull

Andrew Woods ★

Marcus Wust

### 4.32.2 Call-in Details

Dial-in Number: 866-842-5779

Participant Code: 6413798205#

### 4.32.3 Agenda

1. Follow-up to 3/4-5 Tech Team Summit (Scott)
  - a. Goals (as outlined in 2/25/13 call):
    - i. Clarity on roles & expectations: update on overall project trajectory, UVa & DuraSpace contributions, core APTrust staff, contributors from partner institutions
    - ii. Preparation for April 29: review work needed, responsibilities and timelines, outline technical presentation
    - iii. Setting stage for future:
      1. Completion of Phase I - needs clarity:
        - a. business model defines this as through Dec. 2013
        - b. tech model considered Phase I period leading up to demo
        - c. also need to think about when we cross line into operationalization
      2. Preparation for Phase II
        - a. what will business model look like beyond Phase I?
        - b. Could APTrust serve as pass-through service for those interested only in DPN?
2. Update on Advisory Committee call (Donna)
3. Update on forming Policy Group (Greg R)
4. Update on Program Director search (Bradley)

### 4.32.4 Discussion

#### Tech Team Summit Debrief

1. UVa will be trying to hire two new developers
2. Dan Davis will be the sys-admin servers
3. Architecture continues to be cloud-based
4. Do we need to set up new AWS account for UVa?
5. Any core changes in DuraCloud tech-stack will be handled by DuraSpace
6. Looking to DuraSpace where appropriate code updates would fit
7. Interested in including APTrust partners in development
8. Looking for feedback on how best to keep the team clear on plan/progress

#### Preparation for April 29th

1. Donna has sent out invitations to members regarding April 29th meeting
2. Will be breaking out into three groups in the morning

3. Tech demo will be right after lunch
4. Rest of the team needs to lay out plan from April to end of year
5. Community will hopefully set at least part of the agenda

## Setting Stage for Future

1. Phase 1 ends in December 2013
  - The point up until APTrust goes into production
2. Tech model will be fleshed out by April
  - System architecture
  - System workflow
3. Phase 2, was not discussed at Summit
4. DPN
  - There will likely not be a simple pass-through first node service
5. Want to make sure that April meeting is APTrust-focused
  - However, want to address some DPN scenarios that community is interested in
  - What would be the most useful way to discuss DPN with APTrust community?
  - What kinds of things should we be talking about? Donna asked Scott & Andrew W to recommend topics/approaches.
6. Donna and Michele had post-Summit discussion of DuraSpace/UVa partnership
  - A new scope-of-work document is being drafted
  - Scott will need to provide development requests of DuraSpace
7. Could be a good idea to incorporate BagIt into CloudSync

## Advisory Committee Call Update

1. Happened Thursday after Summit
2. Five deans are going to address why their institution is supporting APTrust @ARL
  - MD
  - Notre Dame
  - Columbia
  - Syracuse
  - UNC
3. Martha will present an overview of APTrust planned functionality
4. There will then be a facilitated discussion
  - What are the preservation needs to be addressed
5. This meeting has been slated for one hour
6. There will be a time before the meeting to speak with the deans before they present

7. What should the break-out sessions look like? 1 hour, including reporting back
  - Tech group
  - Content group
  - Policy group
  - Deans: future membership, cost model
8. Will try to record the presentations

## Program Director Search Update

1. Applications are being reviewed for program director
2. Contact Bradley if we know of any good candidates

## General

1. What do we want to use this APTTrust time for
2. Donna does not participate in the tech calls
3. This meeting should focus more on governance and policy
  - Used to make sure the program is moving forward
4. Strategy call still useful every-other-week
  - Scott, Michele, Gregs, others, to be on call
5. Next call will be first week of April

## 4.32.5 Actions

- ☐ [Scott Turnbull](#) to post summit notes to wiki, currently in google-doc
- ☒ [Scott Turnbull](#) and [Andrew Woods](#) come back to Donna on how to address APTTrust community wrt DPN

```
# How is APTTrust different from DPN?
** Compare/Contrast missions
** DPN offers long-term commitment of collection to academy
** DPN provides additional guarantees of redundancy
# What is required to get content into DPN?
** Deposit agreement specifying succession rights
# What does it mean for content to be in DPN?
** No deletes
** Stay tuned for business model
```

- ☐ [Scott Turnbull](#) to come up with tech break-out topics for April

## 4.33 2013-03-26 APTrust Working Group Meeting

---

### 4.33.1 Attendees

- Danny Bernstein ★
- Andrew Curley
- Dan Davis
- ~~Mike Durbin~~
- Greg Jansen
- Scott Turnbull
- Andrew Woods
- Markus Wust

### 4.33.2 Call-in Details

Google+ Hangout link is on the [event](#) page

### 4.33.3 Agenda

1. Pivotal Tracker (Scott)
2. Transforming Meeting from information sharing to discussing work. (Scott)
3. Schedule demo of site with live data? (Scott)
4. Update on position posting. (Scott)
5. More?

### 4.33.4 Previous Actions

- ☒ Danny will get access to Duracloud from Dan Davis.
- ☐ Markus will run a test using SyncTool against AIPs to compare workflows.
- ☐ Dan - Work with Mike to clear up login issue to uva.duracloud.org server.
- ☐ [Michael Durbin](#) to deploy IPS to aptrust.duracloud.org
- ☐ [Michael Durbin](#) to document deployment/configuration of pieces running on aptrust.duracloud.org

### 4.33.5 Discussion

Pivotal Tracker is up! Let us make pivotal tracker the authoritative record of project progress info.

We need to look into these big themes for December:

- \* The APTTrust Bag
- \* Content Model Design - technical group responsibility
- \* Greg Rashke's group will be responsible metadata structures
- \* Admin Tool migration to Hydra

Scott has some concerns about hitting the December goals - we need to cut down on features as much as possible. Quality is more important than Scope.

Story = Who can do what and why? Who = person or process, what = testable functionality, why = purpose

We'll identify any deal breakers and try to limit updates only to blockers.

**Live guided demo of the application: Friday April 5th 2pm Eastern**

Recruiting effort is ongoing.

## 4.33.6 Actions

## 4.34 2013-04-02 APTTrust DPN Node Discussion

---

### 4.34.1 Attendees

- Gregory Jansen
- ~~Tim Simon~~
- ~~Adam Soroka~~
- Scott Turnbull
- Andrew Woods

### 4.34.2 Call-in Details

Google+ Hangout link is on the [event](#) page

### 4.34.3 Agenda

1. Assumptions about APTTrusts role as DPN First Node.
2. Outline of what information we need to properly create versioned DPN bags from non-versioned APTTrust content.



3. Assumptions about APTrust as a DPN Replication node.

## 4.34.4 Discussion

Notes from the Google Hangout below.

APTrust as a DPN Node

## 1. Assumptions about APTrusts role as DPN First Node.

- a. APTrust is currently targeted at preserving whole objects from Partner Repositories and does not version their content.
- b. Internal APTrust Metadata is versioned via Fedora Datastream versioning.
- c. The mission of DPN is to preserve copies of digital content in whatever state and through whatever versions are sent to it. The mission of APTrust is to preserve the current state of digital content from partner institutions.
- d. To support DPN versioning APTrust needs to keep track of files included in digital objects as part of its bag content model to know what differences exist to create differential DPN versioned bag.
- e. Original Partner content is never modified and anything done with it later will be carried out through representational objects. For DPN this means we will try to pass through objects to DPN in as close a representational state as we can to the original digital object.
- f. The fewer levels of serialization and packaging of content from the originating institutions will allow organizations an easier path to restoration and better meets the goals of preservation.
- g. The structure of the APTrust Bag has a high impact on our ability to create DPN bags and high impact on the ability to restore objects at the originating institution.
- h. Do we need some kind of manifest or additional file to guide a future archivist through how to rebuild a Fedora Object? Thoughts, no the serialize datastreams don't need special explanation for restoration, the needed data for restoring them is in the objects.
- i. Need to capture some information about the synctool used to stage content in APTrust as part of its audit trail and some of that information needs to be included in the DPN bag to give future generations some information of how that object was serialized.
- j. All objects written by the appropriate sync tool (ReplicationTaskSuite/Cloudsync) will be put into the data directory of the bag.
- k. We should not create additional levels of serialization for content than is needed to achieve our goals in general.
  - i. Concern?
- l. Open question of whether there would be an advantage of trying to get all content submissions in bag format?
  - i. ?? If we take bags as original content submissions it would mean we don't really have the ability to differentiate DPN bags. we'd have to work around this by processing bags themselves?
  - ii. ?? What level are we ensuring integrity at, the bag level or what is in the bag. Bags don't provide checksums for the entire bag but the speci provides a means to check important objects in the bag.
    1. Should we shoot to preserve the contents of a bag but not the bag itself? Doing so would allow us to modify a bag and not have to go through several levels of serialization, but it would mean they might get bag different whole bags, even though the preserved portions of that bag will still be the same.
    2. i.e. the checksums on objects will always valid in a bag, but the bag might change or grow based on preservation needs.

2. Outline of what information we need to properly create versioned DPN bags from non-versioned APTrust content.
  - a. Need to be able to read what files have been updated in a digital object from one update to another and write those only to a potential DPN bag.

**Next meeting is what to present as DPN as it relates to APTrust for the April Meeting.**

## 4.35 2013-04-09 APTrust Working Group Meeting

---

### 4.35.1 Attendees

- Danny Bernstein
- Andrew Curley
- Dan Davis
- ★
- Mike Durbin
- Greg Jansen
- Scott Turnbull
- Andrew Woods
- Markus Wust

### 4.35.2 Call-in Details

Google+ Hangout link is on the [event](#) page

### 4.35.3 Agenda

1. Demo Server Needs (Scott facilitate w/ Dan, Mike, Danny)
2. Marking remaining tasks in Pivotal (Scott)

### 4.35.4 Previous Actions

- ☐ Markus will run a test using SyncTool against AIPs to compare workflows.
- ☒ Dan - Work with Mike to clear up login issue to uva.duracloud.org server.
- ☒ [Michael Durbin](#) to deploy IPS to aptrust.duracloud.org
- ☒ [Michael Durbin](#) to document deployment/configuration of pieces running on aptrust.duracloud.org

## 4.35.5 Discussion

- Scott: Domain Name?
- Andrew: SSL redirect to investigate, has shown up before, usually a front end problem
- Scott: Do we need SSL for Tomcat, Fedora
- Consensus: likely not
  
- Scott: Should do a clean install to prepare for production, subsequent development
- Scott: Puppet scripts for install
  
- Dan, need to be clear what is needed for the demo and to enable subsequent work
- Mike likes to keep the server deployment as close to production as possible to prepare
  
- Mike: Possible slowness with [aptrust.duracloud.org](http://aptrust.duracloud.org)
  
- Scott: Fedora admin username/password should be changed
- Mike: Only limited or no use of Fedora
- Scott: Want to choose DB for database
  
- Mike: Fedora not really used for demo, kind of in there but can be cut out
  
- Python library used only for DPN integration
- Rabbit MQ only for DPN integration
- Ruby front end later
- Apache not needed
- Maven2 on server, may need module for Ruby build
- We will likely not build on server
- Java 7 is on server
- Scott: Is there a problem with Fedora on Java 7
- Mike, Dan: Only for the JMS module which we are not using
- Dan: I run Fedora on Java 7
- Dan: Going forward Java 6 is EOL so Java 7 is recommended
- Scott: FF is Java 7
- Dan: Java 7 JMS issue for Fedora 3 will be fixed too if we need it
- Mike: I am building on Java 6 for there is not issue either way
  
- Scott: Other stuff, Owner Field UI bug
  
- Mike: Will work on current issues and reassess on Friday
- Mike: Danny is needed to assess the UI issues
- Mike: No showstopper issues but some may features may be mocked for the demo
  
- Mike: We can keep the upload sizes down to improve performance
- Scott: Would like an upload that takes a reasonable amount of time
- Andrew: Traced through steps for DSpace through demo processing

- Consensus: Key know issues for demo were discussed for now

## 4.36 2013-04-16 APTrust Working Group Meeting

---

### 4.36.1 Attendees

- Danny Bernstein
- Andrew Curley
- Dan Davis
- Mike Durbin
- ★
- Greg Jansen
- Scott Turnbull
- Andrew Woods
- Markus Wust

### 4.36.2 Call-in Details

Google+ Hangout link is on the [event](#) page.

### 4.36.3 Agenda

1. Check in on setup and determine for Demo Server. (Dan)
2. Flow of the half hour demo itself, who is speaking and such. (See [Agenda](#)) (Scott)
3. Update on Hiring (Scott)
4. Other Items?

### 4.36.4 Previous Actions

See Pivotal tracker

### 4.36.5 Discussion

- Asked Dan about mirrored version of test server setup
  - Dan performed a snapshot earlier today and is bringing up a new copy on a higher performance VM
  - Scott put a couple issues related to this in Pivotal Tracker
  - Dan expects the new server to be up by tomorrow
  - Will coordinate with Mike when the elastic IP address changes to point to the new server
  - We can have an additional IP address through which we can access the alternate server
  - Question about bringing up DuraCloud on a larger server in order to have ideal time for test
    - we'll scale the content for the demo period, not the server
- Flow of Half-Hour Demo
  - for the technical demo, who should present
    - probably not command-line tools, instead just a quick narrative
    - then show the progress bar and regular parts of the application
    - by the end we should have the content ingested
    - we'll hold off questions until the end
  - Who from DuraSpace is coming for the Demo?
    - Not Danny, likely not Dan, possibly Andrew
- Hiring News
  - Two positions are posted
    - for first two weeks they're internal only
    - Scott has already shopped them around at local and regional forums
  - Possibly we'll schedule hydra training, maybe along with other UVa Hydra needs, and possibly regional institutions (Richmond?)
- Questions
  - Markus asked if we could record the demo and make a recording available.
    - Scott asked if we could record the dry run and edit it into a video we can upload.

## 4.37 2013-05-07 APTrust Working Group Meeting

### 4.37.1 Attendees

- Andrew Curley
- Dan Davis
- Mike Durbin
- Greg Jansen ★
- Scott Turnbull
- Andrew Woods
- Markus Wust

## 4.37.2 Call-in Details

Google+ Hangout link is on the [event](#) page.

## 4.37.3 Agenda

- Feedback on Partner Meeting from each member that was there (Group)
- Server Infrastructure setup and Schedule (Dan)
- Engaging Content group (Scott)
- Next Steps (Scott)

## 4.37.4 Discussion

### Feedback from the partner meeting

Mike

- Went well, demo fine.
- Couple people volunteered to help with FS (Jim Tuttle, Duke) (Deann Buss, Syracuse)
- Tim DiLauro ingest as DSpace, get it out as Fedora (data mapping from DSpace to Fedora)
- community tools for translation
- People may have been hesitant about Fedora 4
- will be interesting to see how well Hydra data model fits our use case

Greg

- engagement with partners will be the ongoing challenge
- engagement will help us separate contributing partners from those less engaged and waiting to see what we do.

Scott

- real problem is going to be the social engineering - we need help from the partners
- Fedora 4 - Robin heard concerns about that.
- Rick Johnson - ND - interested in working on the DEV side - might even hire a FT APT dev.
- Jill Sexton - very interested in getting involved on the content side
- bringing content people into the emerging APT initiative - agreement around foundational operating principles
- managing scope and perception of scope creep

### Server Infrastructure

Dan collected setup information from folks. Went through installation on higher powered Amazon server. Has a handle on most of the install, except RabbitMQ.

- Domain is set up, but need self-signed certs for communication between DPN nodes
- Integration of public SSL cert still needed for APT client/server calls
- For now, let DuraSpace configure and run the DuraCloud instances. (They have an admin console for this.)
- Action Item: Need to have a discussion about data resources. Need to provision individual developers such that they can start and stop the test and dev duracloud instances.
- Action Item: Scott and Dan meet to discuss procedures for system administration, Dan will set up meeting and highlight decisions that need to be made.
- Start with development oriented approach (DevOps), migrating more toward production over time. (Continuous Integration and deployment)
- Scott: we will need to define an operational team at some point, hopefully somewhat separate from the dev team.

### **Engaging Content Group**

- Get a set of people that can dedicate some time, get their feedback on some of the architecture issues.
- their input on audit trail is essential
- work flow issues around updates, including input into DPN updates. (Content replacement vs. fixed durable objects)
- inform our decisions with input from these partners
- We would like partners who contribute their time to understand that they are valued.
- Scott: will meet with Jill and Greg this week to discuss partner contributions from a content perspective

### **Next Steps**

- Need an initiative approach, instead of a project
- how to sustain staffing levels
- shift to a strategic point of view
- how do we manage and engage the community
  - just getting involved with AP Trust becomes a valuable training exercise for their staff
  - new expertise and use cases are brought into the partnership
- become an fedora 4 test site, install it, blog about it, raise any issues with FF
- Do we want to require a bag for all submissions.



## 4.38 2013-05-28 APTrust Working Group Meeting

---

### 4.38.1 Attendees:

- Markus Wust
- Andrew Curley
- Greg Jansen
- Dan Davis
- Jeremy Morgan
- Daniel Bernstein
- Suzanne Preate
- Sarah Timer

### 4.38.2 Discussion

**List of participants in the APTrust project through Sept, matching skills and general availability.**

- Scott Turnbull - UVa: Full time (half time on DPN node of APtrust) Programming, planning and coordinating groups.
- Markus Wust - NCSU: Working on DSpace sample content and packaging with Replication Task Suite.
- Andrew Curley - UVa: Digital Curation Services; Software Developer; Ruby on Rails developer (Hydra); Some portion of time perhaps, still working out details.
- Greg Jansen - UNC: Digital Repository Services Team; Tech Liaison to APTrust partners; Working with Fedora Futures group and will help connect APTrust to Fedora architecture in terms of getting APTrust to work with that; Most of time is on Fedora Futures and that will be used to help connect APTrust.
- Dan Davis - Duraspace: Operations and Development Support; Sys Engineering and Repositories (Fedora) skills. Core messaging and communication infrastructure, Java server side developer. Liaison for Duracloud. ~10 hours a week.
- Jeremy Morgan - Syracuse: PHP developer primarily. Flat file content push and developing workflows around that. Unclear on commitment ~10% of time.
- Danny Bernstein - Duraspace: Duracloud committer; Java, Spring and EE skills. Did first revision of admin tool for APTrust with Mike Durbin. Still working with Duraspace for time and so forth, more information coming.
- Suzanne Preate - Syracuse: Digital Initiatives Library; Project management, workflows and technical metadata. Limited available for June, more available July/Aug
- Sarah Timer - Syracuse: Metadata Librarian; Workflows and interoperability between datasets.

**Goals and subteams for end of June 2013**

1. Getting Replication Task Suite workflow for staging data documented and working out a plan with Duraspace for improving performance of that tool. (Markus, Scott)
2. Create BagIt Profile while with content specialists for storing preservation items in APTrust. (Scott, Jeremy and Possibly Suzanne/Sarah)
3. Document a strawman workflow for packaging Filesystem content and sending to the APTrust staging area. (Greg Jansen, Scott, Jeremy)
4. APTrust can identify when a full Fedora object has landed in Duracloud to know when to kickoff a ingest routine on that item and ensure it doesn't get overwritten before it has been processed. (Danny Bernstein?, Dan Davis).
5. Setting up development environment for non-Duracloud stack. (Dan, Scott)
6. Exploring possibility of using bag storage directly with Fedora 4 and get some feedback about how well that might integrate with our current plan for APTrust. - (Greg)
7. DPN Goals for creating a basic DPN bag for testing and initiating content transfer between DPN nodes through the messaging system. (Scott)

### 4.38.3 Next Steps

- ☒ Scott will follow-up with Jeremy, Suzanne and Sarah from Syracuse for an overview of APTrust as a whole and discuss the BagIt profile and Filesystem packaging needs.
- ☒ Scott will follow-up with each subgroup to get them started, set up a subteam kick off next week or both.

## 4.39 2013-06-04 APTrust Working Group Meeting

---

### 4.39.1 Attendees:

- ~~Markus Wust~~
- ~~Andrew Curley~~
- ~~Greg Jansen~~
- ~~Dan Davis~~
- Jeremy Morgan
- Daniel Bernstein
- ~~Suzanne Preate~~
- ~~Sarah Timer~~
- Michele Kimpton
- Mike Durbin

### 4.39.2 Agenda

1. Update from Subgroups (Scott)

2. Sub Group Coordinators (Scott)
3. Open Repositories Presentation (Scott)
4. Generating Development Goals (Scott)
5. PID issue from community discussion. Want to ask partners, but no clear list for who to ask. (Scott)

### 4.39.3 Updates from Sub Groups

Any recent activity, any need for additional information, resources or clarity?

1. Replication Task Suite
  - a. Requested meeting between Markus, Scott and....
2. BagIt Profile
  - a. First meeting of group is tomorrow. Will have an update for next week.
3. Filesystem Packaging & Staging
  - a. Initial Meeting Today, will update afterwards.
  - b. Archivematica and Curator's Workbench are suggested as possible starting points for processing filesystem content.
4. Development Environment Subteam
  - a. Scott setup new AWS account and setup IAM accounts for administrators.
  - b. Scott is working with Dan Davis to get encrypted email setup in order to pass credentials back and forth for AWS service..
5. Bag Storage Options in Fedora 4
  - a. Work underway, nothing to report this week.
6. Bag Creation Milestone
  - a. Working on initial set of BagIt scripts, code located at <https://github.com/APTrust/bagins>
  - b. Current python libraries inadequate for DPN bags because of custom TagFiles used.
  - c. Speed of bagging libraries questionable for scaling needed.
  - d. Will Benchmark performance to assess usefulness.
  - e. Should be able to create initial bags by next Meeting.
7. DPN Content Transfer Milestone
  - a. Work slow due to high level of APTrust this month and next.
  - b. May lag behind group milestone but making progress.
8. Orienting New Development Team Members
  - a. Hope to reach better clarity on who exactly is in this group and what their availability is.
  - b. Working with Robing Ruggaber at UVa to see if we can schedule a mini-Hydra camp in July.

### 4.39.4 Discussion

- There is no file-size limits on individual files set by Duracloud itself but there is a 5TB limit set by S3 storage.
- The chunking of files by Duracloud happens to enable more reliable transfer of content over HTTPS.

- Danny Bernstein will be a liaison for Duracloud to help us work through APTrust needs and help determine with the Duraspace team.
- Dan Davis may be able to help a bit with profiling information about moving things into and out of Duracloud based on work he's done before.
- Big difference between 1.8x and 3.x line of DSpace in terms of the way AIP are created, this might be worth trying out to see if we can improve performance.
- Presenting at Open Repositories 10 July 9:30-10:30. Suggested participation in the presentation partners to express their view of the value of APTrust how they see their organization using it. .

## 4.39.5 Next Steps & Open Issues

## 4.40 2013-06-18 APTrust Working Group Notes

---

### 4.40.1 Attendees:

- ~~Markus Wust~~
- Andrew Curley
- Greg Jansen
- Dan Davis
- ~~Jeremy Morgan~~
- Daniel Bernstein
- ~~Suzanne Preate~~
- ~~Sarah Timer~~
- Mike Durbin

### 4.40.2 Agenda

1. Update on BagIt scripts (Scott)
2. Update on DPN content transfer (Scott)
3. Update on the BagIt Profile (Sarah or Suzanne)
4. Update on Storage Profiling in Fedora 4 (Greg)
5. Update on Filesystem packaging (Jeremy)
6. Update on bringing in programmers (Scott)
7. Update on OR Presentations (Scott)
8. Update on Server Setup (Dan)
9. Hydra Camp!!! (Scott)

## 4.40.3 Discussion

1. Update on BagIt scripts (Scott)
  - a. Library is still under development.
  - b. VaTech may try out the library to compare performance over large files.
  - c. Finished payload features this week and next week should be able to produce a whole bag.
2. Update on DPN content transfer (Scott)
  - a. Development stalled due to travel.
  - b. Basic workflow tracking has already been added to app.
  - c. Still have to plugin the actual workflow step updates to the message handlers.
3. Update on the BagIt Profile (Sarah or Suzanne)
  - a. Most of team traveling, no update this week.
4. Update on Storage Profiling in Fedora 4 (Greg)
  - a. Greg has been working on some sprints with them around the ModeShape Federation Feature.
  - b. Gathered some promising results from this and there might be some advanced features we want to take advantage of.
  - c. Greg wrote up a good set of notes on Fedora 4 Opportunities in APTTrust in GoogleDocs at [https://docs.google.com/document/d/1\\_omLNsuDbV-zRCc3t3yNbuMXGo01jTV4iFhn4RUfYbM/edit?usp=sharing](https://docs.google.com/document/d/1_omLNsuDbV-zRCc3t3yNbuMXGo01jTV4iFhn4RUfYbM/edit?usp=sharing)
  - d. Fedora 4 supporting good namespacing and nested PIDs.
5. Update on Filesystem packaging (Jeremy)
  - a. No feedback this meeting.
  - b. Greg and Scott will check in with Jeremy on progress and if can help any.
6. Update on bringing in programmers (Scott)
  - a. Talking with several potential partners to get commitment to development.
  - b. No definitive commitment yet but hope to have more information soon.
7. Update on OR Presentations (Scott)
  - a. Scott will be presenting with Rob Carolano giving partner perspective and Greg Jansen on workflow and Fedora 4 perspective as well.
  - b. Need to meet with group.
  - c. Scott also presenting as part of DPN group.
  - d. Thursday 10-11 will be both presentations at OR.
8. Update on Server Setup (Dan)
  - a. Main AWS account setup last week and Dan is setting up new account.
  - b. Established encrypted email connection between APTTrust Admins to pass sensitive information back and forth.
  - c. Dan working on setting up individual developers.
  - d. Will use secret key to authenticate and so forth.

#### 9. Hydra Camp

- a. Scott is working with Robin Ruggaber and UVa to setup a Hydra Camp the first week of August.
- b. Update to community should come out soon.
- c. Will be modified Hydra Camp and asking participants to do Ruby and Rails tutorials before coming. Asking to try Hydra tutorials before coming as well.
- d. APTrust may be able to pay for attendance of anyone from a partner willing to commit them for 50% of time over 3 months sometime in the next 6 months.

## 4.40.4 Next Steps & Open Issues

## 4.41 2013-06-25 APTrust Working Group Meeting

---

### 4.41.1 Attendees:

- ~~Markus Wust~~
- ~~Andrew Curley~~
- Greg Jansen
- Dan Davis
- ~~Jeremy Morgan~~
- Daniel Bernstein
- ~~Suzanne Preate~~
- ~~Sarah Timer~~
- Mike Durbin
- Scott Turnbull

### 4.41.2 Agenda

1. Update on BagIt library (Scott)
2. Update on DPN Content Copy Service (Scott)
3. Update on APTrust Servers (Dan)
4. Update on BagIt profile (Suzanne, Sarah or Jeremy)
5. Update on Filesystem packaging. (Jeremy)
6. Update on Fedora 4 Storage Options (Scott)
7. Check in on OR Presentation (Scott)
8. Update on Connecting with Content Group. (Scott)

### 4.41.3 Discussion

1. Update on BagIt library (Scott)
  - a. Code is progressing well and we should be able to create basic bags for the library by the end of the month.
  - b. Code hosted on Github at <https://github.com/APTrust/bagins>
2. Update on DPN Content Copy Service (Scott)
  - a. No development on this. Likely we will not make our June goal.
3. Update on APTrust Servers (Dan)
  - a. Have a basic AWS account in place and using IAM to create management accounts.
  - b. Each developer will get have their own keypair in the IAM service for them to start-up and shutdown development servers as needed under their key.
  - c. Dan has setup an initial AIM with very basic installs of environmental languages.
  - d. Copy of the server used for the Demo at the Partner meeting is saved as an AMI and can be accessed.
  - e. Will start setting up Puppet Scripts and so forth.
  - f. IAM logins will be added for people deploying services or code as needed.
  - g. Should be putting up Medium Reserved Instance for Demo, Testing and integration testing.
4. Update on BagIt profile (Suzanne, Sarah or Jeremy)
  - a. Still in progress. Travel and other commitments may make this difficult to get thorough.
5. Update on Filesystem packaging. (Jeremy)
  - a. There have been some scheduling difficulties
6. Update on Fedora 4 Storage Options (Scott)
  - a. Checked with Greg Jansen and Adam Soroka for storage scenarios.
  - b. No deal breakers for proposed scenarios so far.
  - c. There is a Bag connector already working for Fedora 4 we will want to test out shortly after getting an install
  - d. Looking to get a Fedora 4 install for development and testing in July and report back via a blog to the community on experiences with that install.
7. Check in on OR Presentation (Scott)
  - a. Will get with Greg later in the week and check with Rob.
  - b. Will post announcement of presentation to blog and community this week.
8. Update on Connecting with Content Group. (Scott)
  - a. Had first meeting this week.
  - b. Subgroups are beginning to ramp up a bit more.

### 4.41.4 Next Steps & Open Issues

- ☐ [Scott Turnbull](#) will post announcement of OR presentation this week.

- ☒ **Scott Turnbull** will get with Greg J and check in with Rob C to confirm presentation details.
- ☒ **Scott Turnbull** should finish off BagIt code this and be able to generate initial test bags.
- ☒ **Scott Turnbull** will send out announcement of hydra camp this week.

## 4.42 2013-07-02 APTrust Working Group Meeting

---

### 4.42.1 Attendees:

- Markus Wust
- Andrew Curley
- Greg Jansen
- Dan Davis
- Jeremy Morgan
- Daniel Bernstein
- Suzanne Preate
- Sarah Timer
- Mike Durbin
- Scott Turnbull

### 4.42.2 Agenda

1. **Report out on each of the Goals from June (Scott, facilitate. each team)**
2. **Summary of Goals for July. (Scott facilitate)**
3. **Update on Open Repository Presentation. (Scott)**

### 4.42.3 Discussion

1. Report out on each of the Goals from June (Scott, facilitate. each team)



- a. Speed up the packaging and staging of DSpace content. Our work so far with ReplicationTaskSuite for staging DSpace content has been promising but the process is slow, on the order of 2 hours for 3.5k PDFs. (Markus Wust [NCSU], Scott Turnbull[APTrust])
  - i. Markus was able to speed up the packaging of content to local storage from 2 hours to about 10 minutes.
  - ii. Next step is to actually sync the content to Duracloud to time the full object creation. Create July goal for this.
- b. Document a BagIt Profile for use in packaging items for storage in APTrust. (Jeremy Morgan [Syracuse], Suzanne Preate[Syracuse], Sarah Theimer[Syracuse], Scott Turnbull,)
  - i. Varying schedules have made coordinating the group difficult in June, will move this goal July.
- c. Test and document a strawman workflow for packaging Filesystem content for staging in APTrust. (Greg Jansen [UNC], Scott Turnbull, Jeremy Morgan)
  - i. Varying schedules have made coordinating the group difficult in June, will move this goal July.
- d. Establish new Amazon Web Services account to host development environment and test environment separate from Duracloud storage for better management. (Dan Davis [Duraspace], Scott Turnbull)
  - i. AWS account has been setup and initial access given to to the administrator as well as a separate developers group created.
  - ii. Creation of initial standard development server deferred to July.
  - iii. Early July goal to create the initial development AMI to enable developers to get started and to perform test installs of Fedora 4.
- e. Work Directly with the Fedora Futures group to test and recommend better storage strategy for production services. - (Greg Jansen)
  - i. Satisfied. Gregs document with an analysis of Fedora 4 here. [https://docs.google.com/document/d/1\\_omLNsuDbV-zRCc3t3yNbuMXGo01jTV4iFhn4RUfYbM/edit?usp=sharing](https://docs.google.com/document/d/1_omLNsuDbV-zRCc3t3yNbuMXGo01jTV4iFhn4RUfYbM/edit?usp=sharing)
- f. Successfully transfer test content between DPN nodes. Using the federated messaging service we've established over the last few months, we should be able demonstrate the ability to initiate and automatically transfer content between DPN nodes.. (Scott Turnbull)
  - i. Unable to satisfy goal in June due to competing priorities. Deferring to July.
- g. Produce initial version of APTrust and DPN bags from test content. Create smalls service to generate BagIt bags from arbitrary content that supports optional tag files and strong cryptographic fixity. (Scott Turnbull)
  - i. Initial BagIt library created and posted to the community on GitHub at <https://github.com/APTrust/bagins>
  - ii. Can create initial basic bags.
  - iii. July goal to create a simple JSON-RPC service around this to connect to processing service.

- h. On Board programmers from partners helping to advance APTrust and set initial programming Goals. (Scott Turnbull)
  - i. Andrew Curley from UVa working with APTrust over next several months on administrative application and API.
  - ii. Jay Chen from VaTech will be working with APTrust over next several months on processing and registering content.
2. Summary of Goals for July. (Scott facilitate)
  - a. Perform a Fedora 4 Install in the AWS space and make an AMI available for development within APTrust and shareable to any APTrust partner who wishes to use it. (Mike Durbin[UVa])
    - i. Goal is dependent on state of Fedora 4 which we should have more information on at OR next week.
    - ii. If installed goes well we will create a new AMI specifically for Fedora 4 use.
  - b. APTrust members can view an initial admin site skeleton with login features, get assigned to institutional roles and edit information about their institution. (Andrew Curley[UVa])
  - c. Create a simple JSON-RPC server around the BagIt Library to create DPN and APTrust bags. (Scott Turnbull)
  - d. Work with DPN nodes to initiate through messaging and successfully transfer content between federation members. (Scott Turnbull)
  - e. Test the full workflow of packaging content from DSpace and move it to the staging instance to compare performance improvements. (Markus Wust[NCSU])
  - f. Establish initial development Amazon Machine Image (AMI) for use in development, testing and deployment of APTrust and experimental services.
  - g. Document a BagIt Profile for use in packaging items for storage in APTrust. (Jeremy Morgan [Syracuse], Suzanne Preate[Syracuse], Sarah Theimer[Syracuse])
  - h. On board additional programmers from partners to advance APTrust mission. (Scott Turnbull)
3. Update on Open Repository Presentation. (Scott)
  - a. Presentation is next Thursday at OR at 10:30

## 4.42.4 Next Steps & Open Issues

 [Scott Turnbull](#) will post announcement of OR presentation this week.

## 4.43 2013-07-16 APTrust Working Group Meeting

### 4.43.1 Attendees:

- ~~Markus Wust~~
- Andrew Curley
- Greg Jansen
- Dan Davis

- ~~Jeremy Morgan~~
- Daniel Bernstein
- ~~Suzanne Preate~~
- ~~Sarah Timer~~
- Mike Durbin
- Scott Turnbull
- Jie Chen

## 4.43.2 Agenda

1. Introduce new group members and connect important subgroups. (Scott)
2. Feedback on BagIt Profile. (Scott facilitate, group)
3. Producing Test Bags (Scott)
4. Schedule Fedora Install w/ Bag Connector. (Scott facilitate. Mike/Dan)
5. Getting Jei, Andrew, Mike access to setup AWS servers (Scott facilitate, Dan)
6. OR Observations (Group)

## 4.43.3 Discussion

1. Observations from OR
  - a. Michele: From Fedora 4 group
    - i. Good opportunity to work together.
    - ii. Going to work with Andrew Woods to do install and ID test cases each month to try out real world scenarios.
2. Introduce new group members and connect important subgroups. (Scott)
  - a. Members gave self intros.
  - b. Welcoming Andrew Curley from UVa working on the Hydra Head, Jie Chen from VaTech working on the ingest processing and bagging, Mike Durbin working on Fedora Installs and configuration, Jill Sexton repping the Content group who is working on several items including the audit metadata.
3. Feedback on BagIt Profile. (Scott facilitate, group)
  - a. Greg: Whole fedora object goes into the data directory of the bag so it can be sequenced and preprocessed later. Dspace AIP contents also go into the data directory.
  - b. Greg: Looks like it will work but will need to test it.
  - c. Uncertain of the ability to connect to S3 and if that is an option right now.
  - d. Greg: Profile seems fine but some concerns around how Fedora 4 connects to bags.
  - e. All Probably need another meeting to talk about storage nuances and such
4. Producing Test Bags (Scott)
  - a. Scott will try to produce some test bags by the end of the week to use to connect Fedora to.
5. Schedule Fedora Install w/ Bag Connector. (Scott facilitate. Mike/Dan)
  - a. Mike, Dan, Greg and Jie will work out a schedule for install sometime this week or next.

6. Getting Jie, Andrew, Mike access to setup AWS servers (Scott facilitate, Dan)
  - a. Dan is ready to help them out with this. His roll is to get them connected but each developer should get to a point of starting and stopping their own isntances.
  - b. Asked to shut down instances after use so there is no un-needed service costs incurred.

## 4.43.4 Next Steps & Open Issues

- ☐ Perform a Fedora 4 install and configure the bag connector [Michael Durbin](#), [Greg Jansen](#), [Daniel Davis](#)
- ☒ Create some initial test bags using the APTrust BagIt Profile @Scott Turnbull

## 4.44 2013-07-23 APTrust Working Group Meeting

### 4.44.1 Attendees:

- Danny Bernstein
- Andrew Curley
- Dan Davis
- Greg Jansen
- Markus Wust

### 4.44.2 Agenda

1. Update on Test Bag Production (Scott)
2. DSpace test bag production (Markus)
3. Bag Storage options (Scott or Jay)
4. Update on Fedora & Bag connector Install (Scott, Mike or Dan)
5. User role management plugin for Hydra (Andrew)
6. BagIt Profile additions (Scott)
7. Help editing the group charter (Scott facilitate, asking for volunteers)
8. Update on Hydra Camp (Scott)

### 4.44.3 Discussion

1. Update on Test Bag Production (Scott)
  - a. Scott published fedora test bags based on UVa content to an S3 bucket and notified Mike, Dan and Jay so they can use the bag connector on it.
  - b. DSpace content will be bagged through the BagIt option in ReplicationTaskSuite (see below)

2. DSpace test bag production (Markus)
  - a. Markus will update via email later. Goal listed below.
3. Bag Storage options (Scott or Jay)
  - a. Jay is investigating the utility of creating an ExternalFileConnector for use with Fedora 4 to connect to cloud based storage.
  - b. Greg will be attending a Fedora sprint in 2 weeks and it would be good if Jay could communicate with Greg any specific questions he may be able to look into during that time.
  - c. Jay added the following update via Email:
  - d.
    - i. After some code reviewing today, I can confirm ExternalBinaryValue is the way modeshape provides for those value not in mode shape repository. Also, [JClouds](#) could be cloud compute/storage wrapper for us. In another word, JCloudsConnector is the general way to store data on the clouds beyond FF repository. Right now, I am: 1) Going through the code of those parts, trying to familiar with them and coding upon them. 2) Doing more research on how to serialize the BagIt info on top of JCloudsConnector, which should abide by the architecture of mode shape.
4. Update on Fedora & Bag connector Install (Scott, Mike or Dan)
5. User role management plugin for Hydra (Andrew)
  - a. Has been looking into how to manage group permissions in Hydra
  - b. Out of the box the groups support is light and done through YAML.
  - c. We want to be able to delegate assignment of institutional roles to institutional admins.
  - d. Working with Boston Public Library to improve a gem they have been developing for this purpose so it will be useful to us, work with Fedora 4 and is useful to the broader community.
  - e. Making a pull request to BPL to get it included in the latest version of the of the gem.
6. BagIt Profile additions (Scott)
  - a. Adding a section to the bag profile around mitigating for malicious tampering of bag content once it has been received.
7. Help editing the group charter (Scott facilitate, asking for volunteers)
  - a. Greg will work with Scott to develop the unwritten sections of the Group Charter to present to the group for feedback soon.
8. Update on Hydra Camp (Scott)
  - a. 23 participants so.
  - b. Will happen week of August 5th.

## 4.44.4 Next Steps & Open Issues

- ☐ Perform a Fedora 4 install and configure the bag connector [Michael Durbin](#), [Greg Jansen](#), [Daniel Davis](#)
- ☐ Followup on sub-group to edit and publish and [Initial Charter Document](#). [Scott Turnbull](#), [Greg Jansen](#)
- ☐ [Scott Turnbull](#) will work to include a recommendation in the [APTrust BagIt Profile](#) to mitigate against malicious tampering of bags maintained by APTrust.
- ☐ [Scott Turnbull](#) to send out Hydra Camp information to attendees this week confirming time and location.
- ☐ [Andrew Curley](#) make a pull request to Boston Public Library's Hydra Role gem to work with Rails 4.

- ☐ Markus Wust will configure ReplicationTaskSuite to export content using the BagIt option and see if there are options for configuring basic APTTrust bag-info.txt fields as part of the bagging.
- ☒ Scott Turnbull will connect Jay and Greg before he goes to the Fedora sprint in 2 weeks so Greg can check on particular issues or answer questions Jay may have as part of his investigation.

## 4.45 2013-07-30 APTTrust Working Group Meeting

---

### 4.45.1 Attendees:

- Danie Bernstein
- Dan Davis
- Greg Jansen
- Jay Chen
- Jill Sexton
- Markus Wust
- Mike Durbin
- Scott Turnbull

### 4.45.2 Agenda

1. Update on User Roles Module (Andrew)
2. Update on Cloud Connector (Jei or Scott)
3. Update on Fedora 4 install w/ BagIt Connector (Mike, Dan)
4. DSpace bag test (Markus)
  - a. Exported DSpace sample content using the BagIt format this time.
  - b. could not configure the bags being generated to conform to APTTrust Bag Profiles. Needed to change bag name, regen the manifest and such.
  - c. This seems to be as much work as it would take to bag it.
5. Update on HydraCamp (Scott)
6. Content Group Update (Jill)

### 4.45.3 Discussion

1. Update on User Roles Module (Andrew)
  - a. Andrew is working with a Hydra gem from Boston Public Library to update it for Rail 4.
  - b. Updates are mostly done. Will associate users with institutions and allow them to be in particular groups per institution.

2. Update on Cloud Connector (Jei or Scott)
  - a. Jei working with Fedora Future group to investigate implementation of an ExternalFileConnector class that would act like the current FileConnector class on Cloud based storage.
  - b. He is targeting the JClouds library to enable this library.
  - c. Looking into decoupling the BagConnector code as a stand alone library so it can be used with either ExternalFileConnector or the FileConnector subclass.
3. Update on Fedora 4 install w/ BagIt Connector (Mike, Dan)
  - a. Targeting a Fedora 4 install with bags on Elastic Storage
  - b. Needed to investigate how bags are represented to backwards engineer a Hydra Model based on the representation of bags in Fedora.
  - c. Having this by the end of the week will allow us to present examples and use cases to instructors at the Hydra Camp for questions.
  - d. Mike needs to have his key generated so he can spin up a server and Dan will help him through this.
4. DSpace bag test (Markus)
  - a. Results of ReplicationTaskSuite bag production are that bags do not have a high level of customization.
  - b. Produced bags would be renamed, an additional checksum done as it only offers md5, scripts needed to add info to the bag-info.txt tag file.
  - c. This is about as much work as generated new bags from AIPs.
  - d. Suggest keeping DSpace exports as AIPs since it is no more work to bag them later, and doing so would keep the logic for bagging in the same place for all workflows (DSpace or Fedora)
5. Update on HydraCamp (Scott)
  - a. Camp on track for next week.
  - b. Monday is the big Q/A session so please get any use cases or examples in by then to ask the instructors about APTrust specific use cases.
6. Content Group Updated
  - a. First big meeting last week.
  - b. Created a Google+ community to communicate.
  - c. Several folks working on the TDR issues related to preservation requirements.
  - d. Working on preservation audit metadata. Using Archivematica PREMIS events as basis.

## 4.45.4 Next Steps & Open Issues

- ☒ Perform a Fedora 4 install and configure the bag connector [Michael Durbin](#), [Greg Jansen](#), [Daniel Davis](#)
- ☐ Follow-up on sub-group to edit and publish and [Initial Charter Document](#). [Scott Turnbull](#), [Greg Jansen](#)
- ☐ [Scott Turnbull](#) will work to include a recommendation in the [APTrust BagIt Profile](#) to mitigate against malicious tampering of bags maintained by APTrust.
- ☒ [Scott Turnbull](#) to send out Hydra Camp information to attendees this week confirming time and location.
- ☐ [Andrew Curley](#) make a pull request to Boston Public Library's Hydra Role gem to work with Rails 4.
- ☒ [Markus Wust](#) will configure ReplicationTaskSuite to export content using the BagIt option and see if there are options for configuring basic APTrust bag-info.txt fields as part of the bagging.

## 4.46 2013-08-13 APTrust Working Group Meeting

---

### 4.46.1 Attendees:

- Dan Davis
- Greg Jansen
- Jay Chen
- Markus Wust
- Scott Turnbull

### 4.46.2 Agenda

1. Update on Bag Object expressed through Fedora (Scott)
2. Use of Travis for CI (Andrew)
3. Update on format alerts service for file-mimetypes (Scott)
4. Next Steps (Scott)

### 4.46.3 Discussion

1. Update on Bag Object expressed through Fedora (Scott)
  - a. During Hydracamp last week we created a Bag Hydra Model and associated Hydra Datastreams.
  - b. Datastreams on bags currently represent PREMIS Events and Metadata about files in the bag.
2. Use of Travis for CI (Andrew)
  - a. Improved our Test Driven Development infrastructure last week.
  - b. Connected GitHub repositories to Travis Continuous Integration Server and Code Coverages.
  - c. Links to those services are available from the GitHub Repository on the front page of the repository.
  - d. Master and Develop are both being automatically built and tested via Travis.
  - e. Documentation generated for those services over the next few weeks.
3. Update on format alerts service for file-mimetypes (Scott)
  - a. Scholars Lab developers have been building a JSON file with file alert data from PRONOM and UFDR databases
  - b. Will use at some opportune point in APTrust to register file alerts based on mimetypes of files being preserved.
4. Content group (Scott)
  - a. Meeting later in the week, will send them the initial fields created for the PREMIS Events Datastream to help shape that discussion.



#### 5. Next Steps (Scott)

- a. User Stories for Admin interface are in the PivotalTracker site.
- b. Continue to mature the server infrastructure.
- c. Script to attach bags to description objects and begin to get the appropriate metadata files index in Solr.

### 4.46.4 Next Steps & Open Issues

- ☐ Follow-up on sub-group to edit and publish and [Initial Charter Document](#). [Scott Turnbull](#), [Greg Jansen](#)
- ☐ [Scott Turnbull](#) will work to include a recommendation in the [APTrust BagIt Profile](#) to mitigate against malicious tampering of bags maintained by APTrust.
- ☐ [Andrew Curley](#) make a pull request to Boston Public Library's Hydra Role gem to work with Rails 4.
- ☐ [Scott Turnbull](#) will setup a meeting with [Danny Bernstein](#) to talk through Duracloud scenarios for APTrust.

## 4.47 2013-08-20 APTrust Working Group Meeting

---

### 4.47.1 Attendees:

- Dan Davis
- Greg Jansen
- Jay Chen
- Markus Wust
- Scott Turnbull

### 4.47.2 Agenda

1. Update from Content Group (Jill Sexton)
2. Update on APTrust Open Positions (Scott)
3. Update on Bagging Service. (Scott)
4. Update on RabbitMQ Federation Plugin (Scott)
5. Update on Fedora Cloud Connector (Scott)
6. Update on Admin Interface (Andrew)

## 4.47.3 Discussion

1. Update from Content Group (Jill Sexton)
  - a. Content group has met and looked at the materials Scott has asked them to consult. Looked at the TDR recommendations.
  - b. Will make recommendations to contributing institutions, like virus checking should be done locally.
  - c. Will recommend fairly basic use cases.
  - d. Scott mentions that we will be recording PREMIS events using RDF and adhering to existing ontologies.
  - e. Jill will send a document in the next couple of days. After that she wants feedback to resubmit back to the Content Group
    - i. Scott: What are the expectations of material going back and forth to/from DPN?
    - ii. Jill: It should be something query-able through the API.
2. Update on APTrust Open Positions (Scott)
  - a. Currently two open advertised positions for APTrust: Rails Engineer and Preservation Engineer.
  - b. No good candidates for Preservation Engineer
  - c. Good candidate for Rails Engineer.
  - d. Suzanne Thorin is starting to ramp up on her APTrust work. Her first big missions will be to solidify the business model. She needs to be appraised of what engineering decisions will affect the cost model.
  - e. Virginia Tech will be coming on as APTrust and have posted a systems administrator. This person will come on part time to help with managing servers.

3. Update on Bagging Service. (Scott)
  - a. Working with Wayne/Eric from UVa Scholars Lab.
    - i. Additionally working on a format alert mechanism.
  - b. They have a narrative for functional requirements, summarizing below
    - i. An automatic service can listen for new files landing in the APTTrust Staging area so it can initiate bagging when a full object has landed.
    - ii. An automatic service will parse the PID from filenames landing in the staging area to uniquely and consistently identify bag during initiation of the process.
    - iii. An asynchronous work process can keep track of and report on the progress of items being bagged so institutional administrators can view information about the status of items being processed.
    - iv. A command line admin can halt the initiation of new bags being processed so operations can be gracefully shutdown or halted for system configuration changes.
    - v. An automated service can check completed bags against the internal object manifest to ensure all items have been copied correctly and the complete object is preserved in the bag.
    - vi. An automated service can parse metadata about digital objects in a bag to register metadata about object relationships, title and other important metadata to improve object searching, link related items in the admin interface and provide better reporting.
    - vii. An automated process can update Fedora with appropriate metadata processed during bagging of a digital object so it can manage the preservation of those objects and report important preservation events.
4. Update on RabbitMQ Federation Plugin (Scott)
  - a. AMI established in Amazon web service. Scott wants to install and configure to work with DPN
5. Update on Fedora Cloud Connector (Scott)
  - a. Jay is working on code to plug into Fedora 4 that will plug into multiple cloud providers. Soon we may be able to see bags in Fedora in the next few weeks.
6. Update on Admin Interface (Andrew)
  - a. Working on how to serialize Bags into RDF and mastering the ActiveFedora syntax.
  - b. Currently determining what information from Bag objects need to be indexed into the Description Object since that is going to be the level of discover.y
  - c. Need to follow up on the BPL gem pull request.

## 4.47.4 Next Steps & Open Issues

- ☐ Follow-up on sub-group to edit and publish and [Initial Charter Document](#). [Scott Turnbull](#), [Greg Jansen](#)
- ☒ [Scott Turnbull](#) will work to include a recommendation in the [APTrust BagIt Profile](#) to mitigate against malicious tampering of bags maintained by APTTrust.
- ☒ [Andrew Curley](#) make a pull request to Boston Public Library's Hydra Role gem to work with Rails 4.
- ☐ [Scott Turnbull](#) will setup a meeting with [Danny Bernstein](#) to talk through Duracloud scenarios for APTTrust.

## 4.48 2013-08-27 APTrust Working Group Meeting

---

### 4.48.1 Attendees:

- Markus Wust
- Dan Davis
- Danny Bernstein
- Andrew Curley
- Jay Chen
- Greg Jansen

### 4.48.2 Agenda

1. Update on RabbitMQ and Fedreation Plugin install (Scott)
  - a. Next steps in securing it, vhost plan.
  - b. Possibility of RabbitMQ Shovel to JMS on Duracloud?
2. Update on Admin interface coding. (Andrew)
  - a. Steps toward defining and API (Scott)
  - b. Update on Internal RDF structure for datastreams (Scott)
3. Request for help understanding Fedora 4 scenarios a bit better (Scott lead, question for group)
  - a. I'm getting conflicting feedback on whether it works with cloud storage or not?
  - b. What considerations do I need to make on our end with Hydra?
  - c. Unsure what features of Fedora I'll be able to use in a cloud based environment?
4. Preparing for agenda items for Oct 8th Partner Meeting (Scott)
  - a. Topics I am asking to be on the agenda.
  - b. Other topics to be added?

## 4.48.3 Discussion

1. Update on RabbitMQ and Fedreation Plugin install (Scott)
  - a. Next steps in securing it, vhost plan.
    - i. We will eventually want a vhost for DPN and another for any APTTrust MQ stuff. This will be done later.
    - ii. DPN requires the vanilla install with no Vhost, standard Ports default logins for now.
    - iii. Will eventually want to IP restrict access. Should setup port restriction web front end.
  - b. Possibility of RabbitMQ Shovel to JMS on Duracloud?
    - i. Looking into possibility of getting notifications from Duracloud by connecting RabbitMQ via a shovel plugin.
      1. Should allow us to connect to other codebases that use AMQP.
    - ii. This will allow us to connect notifications to non-java based code for task queues to process content.
      1. Looking for this as a solution to notifying bagging service when new files arrive.
2. Update on Admin interface coding. (Andrew)
  - a. Steps toward defining and API (Scott)
    - i. Will work with Andrew to list current API calls based on routes.
    - ii. Want to limit the exposed API calls to only those needed.
  - b. Update on Internal RDF structure for datastreams (Scott)
    - i. All datastreams right now use the RDFDatastream class in Hydra.
    - ii. Are serialized to NTriples.
    - iii. Use standard ontologies.
    - iv. Have datastreams for PREMIS events and one to represent bags.
3. Request for help understanding Fedora 4 scenarios a bit better (Scott lead, question for group)
  - a. I'm getting conflicting feedback on whether it works with cloud storage or not?
    - i. Unsure of what I can and can't do with cloud storage. I know it will work in some respect but I'm unsure what features are or aren't available.
  - b. What considerations do I need to make on our end with Hydra?
    - i. Will F4 honor the current API contract and will Hydra work or do we need to make more modifications?
  - c. Unsure what features of Fedora I'll be able to use in a cloud based environment?
    - i. Currently we're mostly targeting F4 as just a metadata store and all the functionality is controlled by processes outside fedora. Open question of how much we can rely on F4 to manage files, checksums and the like in a cloud environment or do we need to maintain those as external services.

4. Preparing for agenda items for Oct 8th Partner Meeting (Scott)
  - a. Topics I am asking to be on the agenda.
    - i. \* Presentation on progress so far.
    - ii. \* What is sufficient for geographic diversity (present cost models for various options) Do we have to use different AWS regions (one cost impact) or is it sufficient to store using the internal diversity with S3 and another copy in same region Glacier to mitigate against a single tech stack failure?
    - iii. Packaging requirements and what do we want out of packaging items. Is it Mirroring a repository or do we want to abstract away from specific repository implementations.
    - iv. Audit Trails and tracking, present this and get feedback.
  - b. Other topics to be added?
    - i. Asked to be emailed additional suggestions.

## 4.48.4 Next Steps & Open Issues

- ☐ Follow-up on sub-group to edit and publish and [Initial Charter Document](#). [Scott Turnbull](#), [Greg Jansen](#)
- ☒ [Scott Turnbull](#) will work to include a recommendation in the [APTrust BagIt Profile](#) to mitigate against malicious tampering of bags maintained by APTrust.
- ☒ [Andrew Curley](#) make a pull request to Boston Public Library's Hydra Role gem to work with Rails 4.
- ☒ [Scott Turnbull](#) will setup a meeting with [Danny Bernstein](#) to talk through Duracloud scenarios for APTrust. Friday at 1 EST

## 4.49 2013-09-03 APTrust Working Group Meeting

### 4.49.1 Attendees:

### 4.49.2 Agenda

1. Update on Datastream format. (Scott)
2. Requirements and Scheduling to get Medium Reserved Instance Setup (Dan)
3. Update on task queue setup (Scott)
4. Update on Rails Engineer (Scott)
5. Questions on Partner Update (Group)

## 4.49.3 Discussion

1. Update on Datastream format. (Scott)
  - a. See example format below for our RDF Datastreams.
  - b. Taking the position that linked data has a distinct advantage for APTrust's use case.
  - c. Need to discuss more the assumptions on packaging. With the point of view of future services and aggregate repository there could be an advantage of being able to reference items in a bag by it's URI. Linked data would provide a great advantage if we had reliable URIs
2. Requirements and Scheduling to get Medium Reserved Instance Setup (Dan)
  - a. Dan delayed last week a bit.
  - b. Doesn't need additional help right now, working forward.
3. Update on task queue setup (Scott)
  - a. Scott working with Eric from Scholar's Lab to get rudimentary task queue.
  - b. Looking at resque for our internal task queue.
  - c. Doing performance profiling on pure ruby services, vs decoupled services being called from the work queue. Speed benchmarks were established last week. You can see the result here <https://gist.github.com/erochest/6344191>
  - d. Next we'll produce similar results for memory next week.
4. Update on Rails Engineer (Scott)
  - a. Details shared verbally. Inappropriate to post in the public wiki.
5. Questions on Partner Update (Group)
  - a. No specific questions from the group.

Example RDF Output for bagManifest Datastream. (Note, this is a sample with only partial bag, 2 files to give an idea of the structure)

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:ns0="http://download.org/Code/RDF/File_Properties/schema#" xmlns:ns1="http://www.w3.org/2006/http#" xmlns:ns2="http://purl.org/dc/terms/" xmlns:ns3="http://xmlns.com/wordnet/1.6/">
  <rdf:Description rdf:about="info:fedora/___DO_NOT_USE___">
    <ns0:File>
      <rdf:Description rdf:about="https://s3.amazonaws.com/aptrust_test_bags/uva_uva_lib_1229365/bag-info.txt">
        <ns0:checksum rdf:nodeID="g70216030556320"/>
        <ns0:created>2013-09-03 14:59:12 -0400</ns0:created>
        <ns0:format>text/plain</ns0:format>
        <ns0:modified>2013-09-03 14:59:12 -0400</ns0:modified>
        <ns0:size rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">3456</ns0:size>
        <ns1:absoluteURI>https://s3.amazonaws.com/aptrust_test_bags/uva_uva_lib_1229365/bag-info.txt</ns1:absoluteURI>
      </rdf:Description>
    </ns0:File>
    <ns0:File>
      <rdf:Description rdf:about="https://s3.amazonaws.com/aptrust_test_bags/uva_uva_lib_1229365/bagit.txt">
        <ns0:checksum rdf:nodeID="g70216010221000"/>
        <ns0:created>2013-09-03 14:59:12 -0400</ns0:created>
        <ns0:format>text/plain</ns0:format>
        <ns0:modified>2013-09-03 14:59:12 -0400</ns0:modified>
        <ns0:size rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">3456</ns0:size>
        <ns1:absoluteURI>https://s3.amazonaws.com/aptrust_test_bags/uva_uva_lib_1229365/bagit.txt</ns1:absoluteURI>
      </rdf:Description>
    </ns0:File>
    <ns2:title>uva_uva_lib_1229365</ns2:title>
    <ns1:absoluteURI>https://s3.amazonaws.com/aptrust_test_bags/uva_uva_lib_1229365</ns1:absoluteURI>
  </rdf:Description>
  <ns0:Checksum rdf:nodeID="g70216010221000">
    <ns0:checksumValue>ada799b7e0f1b7a1dc86d4e99df4b1f4</ns0:checksumValue>
    <ns2:created>2013-09-03 14:59:12 -0400</ns2:created>
    <ns3:Algorithm>md5</ns3:Algorithm>
  </ns0:Checksum>
  <ns0:Checksum rdf:nodeID="g70216030556320">
    <ns0:checksumValue>cbc799b7e0f1b7a1dc86d4e99df4b2f6</ns0:checksumValue>
    <ns2:created>2013-09-03 14:59:12 -0400</ns2:created>
    <ns3:Algorithm>md5</ns3:Algorithm>
  </ns0:Checksum>
</rdf:RDF>
```

## 4.49.4 Next Steps & Open Issues

- ☐ Follow-up on sub-group to edit and publish and [Initial Charter Document](#). [Scott Turnbull](#), [Greg Jansen](#)



## 4.50 2013-09-10 APTrust Working Group Meeting

---

### 4.50.1 Attendees:

### 4.50.2 Agenda

1. Fedora 4 installation tests. (Scott)
2. DPN Setup for servers (Scott)
3. Confirm Web Server needs (Group/Scott)
4. Next steps for interface (Andrew/Scott)

### 4.50.3 Discussion

1. Fedora 4 installation tests. (Scott)
  - a. Dan: Suggests we deploy it to an on demand instance as it's unstable.
  - b. We don't need this up all the time so on demand instance is fine.
  - c. We can deploy and update as needed to the on demand instance.
  - d. Using this to test against our code base and get a better picture of how cloud storage works with the Fedora 4.
  - e. Work with Mike Durbin to ask him to do another install of Fedora 4 but screen cast it with Dan and I to watch and record the session.
  - f. REMINDER: Setup notes should be kept on the wiki.
2. DPN Setup for servers (Scott)
  - a. RabbitMQ working and is setup.
  - b. Setup SSH key based authentication.
3. Confirm Web Server needs (Group/Scott)
  - a. Mod\_wsgi will be tagged to python3 as the compiler version
  - b. Passenger will be the rails engine under Apache.
4. Next steps for Andrew interface (Andrew/Scott)
  - a. Focus on Individual Item view and displaying detailed information about specific bags.
  - b. Come up with a way to display a basic health check results about a bit.
    - i. Should only be based on the summary of the latest health checks of each file in the bag.
    - ii. Good checks should just read okay.
    - iii. Failed checks should indicate it and provide a way to filter to bags with failed checks.
      1. Individual Item detail views should contain the full information about which files in a bag passed or failed their health check.

## 4.50.4 Next Steps

- ☐ [Scott Turnbull](#) will confirm the installation of RabbitMQ on test.aptrust.org and that it connects to the DPN Federation.
- ☐ [Scott Turnbull](#) will followup with some information to [Daniel Davis](#) about getting rsync setup to test direct DPN replication.
- ☐ [Andrew Curley](#) Will look into Puppet/Capistrano deployment for Rails Apps and document a recommendation.

## 4.51 2013-09-17 APTrust Working Group Meeting

---

### 4.51.1 Attendees:

- [Scott Turnbull](#)
- [Daniel Davis](#)
- Jay Chen
- [Greg Jansen](#)
- [Danny Bernstein](#)
- Suzanne Thorin

### 4.51.2 Agenda

1. Update on Admin Interface and overview of current state (Andrew)
2. Update on Task Queue work w/ Scholars Lab (Scott)
3. Update on BagIt Library changes this week (Scott)
4. Server updates and latest (Dan)

### 4.51.3 Discussion

1. Update on Admin Interface and overview of current state (Andrew)
  - a. Andrew will update with a summary of where the interface is next week.
  - b. Basically indexing, searching and basic datastructures are in place.
  - c. Need to focus now on packaging, storage layer and task queue to enable next set of interface changes.

2. Update on Task Queue work w/ Scholars Lab (Scott)
  - a. Working with Eric Rochester from Scholars lab to test out Resque Task Queue for Ruby.  
<http://rubydoc.info/gems/resque/1.25.0/frames>
  - b. Will connect it to Rails front end to leverage permissions used already and ensure users only see processing of content from their Institution.
  - c. Investigating GoWorker (<http://www.goworker.org/>) that allows you to connect GoLanguage processes to a Requeu Task Queue and initiate them from Ruby as a possible connector.
  - d. Eric working on connecting Reque to rails app for display while Scott is working in additions to Golang BagIt library. They will work together next week to connect them with GoWorker.
3. Update on BagIt Library changes this week (Scott)
  - a. Scott will add ability to open bags to the library.
  - b. Scott will add ability to check bags against a JSON file that indicates the required bag profile.
4. Server updates and latest (Dan)
  - a. Prepping Rsync test for DPN.
  - b. Has been sent the initial Public Keys for other DPN members except TDR, which we should get today.
  - c. Will try to test next week.

## 4.51.4 Next Steps

- ☒ [Scott Turnbull](#) will confirm the installation of RabbitMQ on test.aptrust.org and that it connects to the DPN Federation.
- ☒ [Scott Turnbull](#) will followup with some information to [Daniel Davis](#) about getting rsync setup to test direct DPN replication.
- ☐ [Andrew Curley](#) Will look into Puppet/Capistrano deployment for Rails Apps and document a recommendation. (Cancelled?)
- ☐ [Scott Turnbull](#) to get with Greg J for overview of arbitrary properties on Fedora Nodes.

## 4.52 2013-09-23 APTrust Working Group Meeting

### 4.52.1 Attendees:

- [Scott Turnbull](#)
- [Daniel Davis](#)
- Jay Chen
- [Greg Jansen](#)
- [Danny Bernstein](#)
- Linda Newman

## 4.52.2 Agenda

1. Overview of presentation for the partner meeting. (Scott)
2. Quick run-through of the Demo App and what we'll do between now and then. (Scott)
3. Break out session ideas so far (Greg/Scott)
4. Other questions on Fedora 4 for partner presentation (Greg/Scott)

## 4.52.3 Discussion

Scotts Notes: I was presenting so these notes are light and after the fact.

1. Overview of presentation for the partner meeting. (Scott)
  - a. Shared draft document schedule.
  - b. For Scott's segment:
    - i. Add how this sets us up for services by the end.
    - ii. Give some examples of concrete services we could build.
2. Quick run-through of the Demo App and what we'll do between now and then. (Scott)
  - a. Showed app as it exists.
  - b. Will try to add display of audit datastreams and such.
  - c. Will try to add stubbed displays for items being processed.
  - d. Will try to add display of audit items related to specific files in a bag from the bag item summary view.
3. Break out session ideas so far (Greg/Scott)
  - a. Covered intent of feedback groups.
  - b. Hope to get specific sub-groups identified for BagIt production.
4. Other questions on Fedora 4 for partner presentation (Greg/Scott)
  - a. No further questions.

## 4.52.4 Next Steps

- ☒ [Scott Turnbull](#) will confirm the installation of RabbitMQ on test.aptrust.org and that it connects to the DPN Federation.
- ☒ [Scott Turnbull](#) will followup with some information to [Daniel Davis](#) about getting rsync setup to test direct DPN replication.
- ☒ [Andrew Curley](#) Will look into Puppet/Capistrano deployment for Rails Apps and document a recommendation. (Cancelled. Andrew departing for new job.)
- ☐ [Scott Turnbull](#) to get with Greg J for overview of arbitrary properties on Fedora Nodes. (On hold until after Oct 8th Meeting)

## 5 AP Trust Package Submission Metadata

For packages of content submitted to APTTrust, we require a small amount of metadata for the purpose of facilitating very basic administrative functions. Initial conversations assumed a very simplistic approach that lacked complex structured materials, but it soon became clear that many of the participating institutions already had highly structured data (in fedora or other management systems) and our submission metadata needs evolved.

### 5.1 Phase 1 Functional Requirements

The metadata must support:

1. Identification of materials (so that when you submit something, you can find and export it later, even if /before richer metadata embedded in your content hasn't been processed/indexed)
2. Flagging of content for inclusion into DPN
3. Access controls

It would be ideal for phase I, to have the providing of this metadata be as easy as possible for partner institutions.

The table below lists package metadata needed for the basic administrative functions that must be supported with possible places where that information could be obtained.

Bit of Metadata	Need to support...	how it should be conveyed at ingest time	where it should be stored on objects in aggregation repository
institution	access controls	implied by DuraCloud storage space	hasOwningInstitution RDF triple on the <a href="#">APTrust package object</a>
title	identification (optional)	?	hasTitle RDF triple on the <a href="#">APTrust package object</a>
package identifier	identification	?	URI of the <a href="#">APTrust package object</a>
dpn-bound	DPN inclusion	?	isDPNBound RDF triple on the <a href="#">APTrust package object</a>
rights information	access controls (APTrust and DPN)	?	hasAccessControlPolicy RDF triple on the <a href="#">APTrust package object</a>
objects in package	identification	?	includesResource RDF triple on the <a href="#">APTrust package object</a>

Originally "File type" and "Bag Name" were mentioned, but none of these apply to the management requirements for packages in phase 1 and have been removed for now.

## 5.2 Outstanding Questions

---

1. We've settled on something called a "package" that represents the unit of management for the admin interface, but some required metadata may need to be applied at different levels of granularity.
  - a. content type type will be needed in phase 2 and will be at a much lower level (fedora datastream)
  - b. rights management **may** need to be at a tighter level (possibly objects contain metadata that is public with content that is rights-restricted)
  - c. will users want to delete one "object" from a "package"?
2. Should we prescribe and enforce package sizes more rigorously to avoid allowing users to create a management nightmare? Is that even possible given the widely varying use cases?

## 5.3 APTrust Package Object

---

APTrust Package objects are defined groups of objects that represent a single "package". Metadata and relationships to the included resources all are stored as part of the package object so therefore an object need not be modified in any way to have this grouping and metadata applied to it. Whether these package objects exist before ingest into AP Trust or not is an open question.

### 5.3.1 Package Object Creation

#### Content From Fedora

Initially it was envisioned that institutions could prepare their repository for APTrust ingest by creating package objects in their local repository. Those objects and their relationships would be queried to determine which objects should be ingested into APTrust and provide the metadata required.



This model is being reconsidered in favor of something that:

- involves less work for the submitting institution
- requires no changes (object creation) in the local repository

#### Content from DSpace

*TBD*

## 5.3.2 Relationships

The APTrust Package objects that are retained in the APTrust aggregation repository will be related to objects (either imported from fedora repositories or created to wrap non-fedora content) using the "includesResource" predicate below. Assertions about metadata for the package will also live in the RELS-EXT datastream for the package object and use the predicate names defined for the project.

aptrust-rdf.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF base="http://www.aptrust.org/relationships#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"

  <rdf:Property rdf:ID="includesResource">
    <rdfs:label xml:lang="en-US">Includes Resource</rdfs:label>
    <rdfs:comment xml:lang="en-US">A property that represents a link between an APTrust package
      object and another resource that is considered to be included in that package.</rdfs:
comment>
  </rdf:Property>

  <rdf:Property rdf:ID="isDPNBound">
    <rdfs:label xml:lang="en-US">Is DPN-bound</rdfs:label>
    <rdfs:comment xml:lang="en-US">A property that represents a link between an APTrust package
      and the "true" or "false" literal to indicate whether the resources in that package are
      flagged for inclusion in the Digital Preservation Network.</rdfs:comment>
  </rdf:Property>

  <rdf:Property rdf:ID="hasOwningInstitution">
    <rdfs:label xml:lang="en-US">Has Owning Institution</rdfs:label>
    <rdfs:comment xml:lang="en-US">A property that represents a link between an APTrust package
      and the URI or literal for the owning organization.</rdfs:comment>
  </rdf:Property>

  <rdf:Property rdf:ID="hasTitle">
    <rdfs:label xml:lang="en-US">Has Title</rdfs:label>
    <rdfs:comment xml:lang="en-US">A property that represents a link between an APTrust package
      and the literal representing its human-readable title.</rdfs:comment>
  </rdf:Property>

  <rdf:Property rdf:ID="hasAccessControlPolicy">
    <rdfs:label xml:lang="en-US">Has Access Control Policy</rdfs:label>
    <rdfs:comment xml:lang="en-US">A property that represents a link between an APTrust package
      and either a URI representing the policy or the literal "world", "institution" or
      "private" for the simple preset policies.</rdfs:comment>
  </rdf:Property>

  <!--
    Something like what follows could eventually be useful, in that it would require fewer
    redundant RDF assertions and allow for simple inclusion of sections of an RDF graph.
```

```

<rdf:Property rdf:ID="packageFollowsRelationship">
  <rdfs:label xml:lang="en-US">Package follows Relationship</rdfs:label>
  <rdfs:comment xml:lang="en-US">A property that represents a link between an APTrust package
    and a predicate that itself should be considered to link (and tie in) objects that are
    related to included resources. For example, if this references an "hasPart"
    relationship, then any objects that are the object of an "hasPart" relationship with
any
    subject already included as a resource in this package it is itself considered a part
of
    this package.</rdfs:comment>
</rdf:Property>

<rdf:Property rdf:ID="packageFollowsInverseRelationship">
  <rdfs:label xml:lang="en-US">Package follows Inverse Relationship</rdfs:label>
  <rdfs:comment xml:lang="en-US">A property that represents a link between an APTrust package
    and a predicate that itself should be considered to link (and tie in) objects that
    relate to included resources. For example, if this references an "isPartOf"
    relationship, then any objects that are the subject of an "isPartOf" relationship with
    any object already included as a resource in this package it is itself considered a
part
    of this package.</rdfs:comment>
</rdf:Property>
-->

</rdf:RDF>

```



## 6 Architecture

---

### 6.1 This page is currently being edited to reflect the current development and workflow strawman scenario for group feedback.

---

### 6.2 Current AP Trust Functional Requirements

---

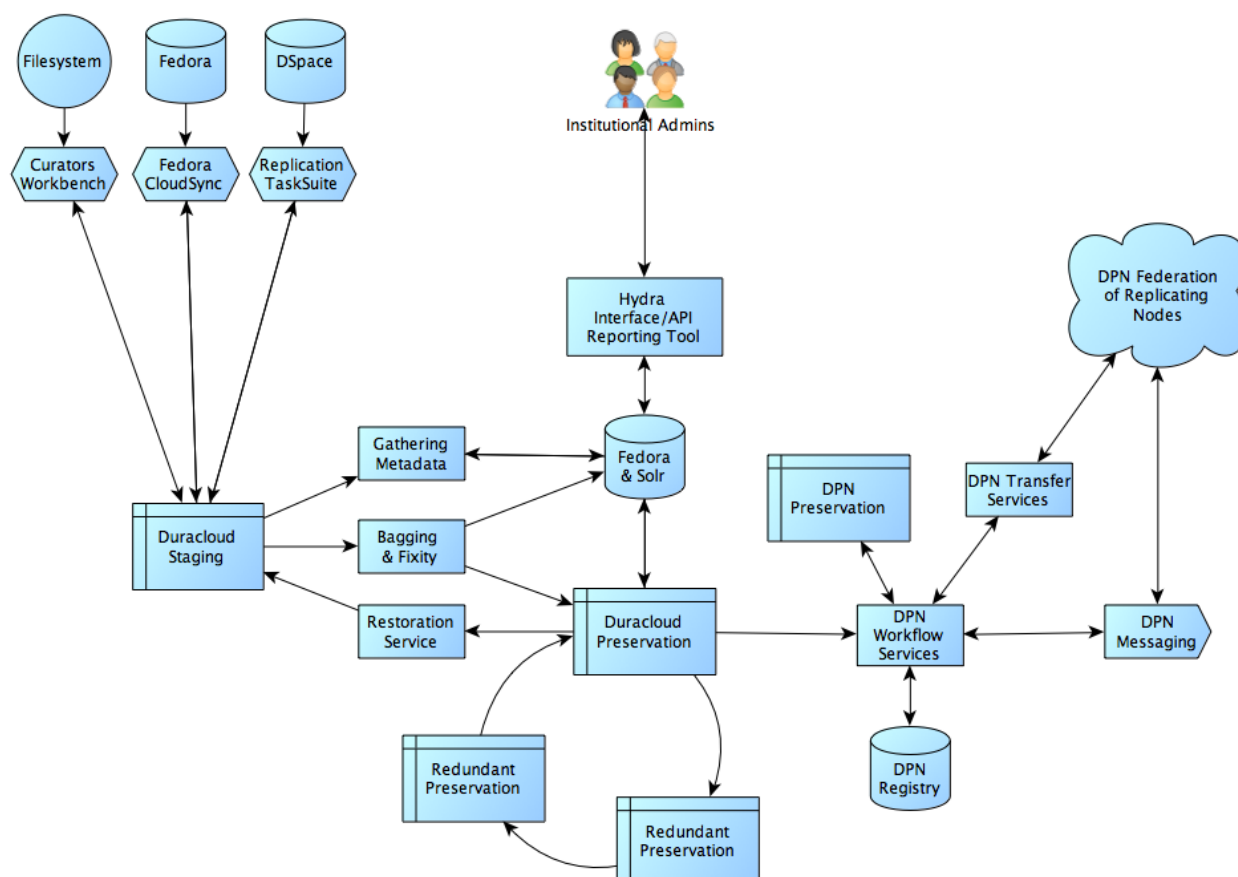
- User submits digital object and can retrieve that exact same object in return
- Every submitted package must have a unique and persistent ID within each staging area.
- The system must accept content originating in Fedora, DSpace or on filesystem in a format agnostic way.
- Content may be Updated freely by the original institution.
- Content may be deleted from APTrust with a corresponding 'tombstone' left for tracking and auditing. Items sent to DPN will never be deleted.

### 6.3 Overview of Flow of Digital Objects

---

At a high level, the intent is to let Partners serialize and push existing managed filesystem or repository content to the APTrust preservation repository. Content will be serialized and copied from a Partner to a staging area, processing scripts will detect the arrival of new content for processing and serialize it using BagIt to create a preservation copy and store metadata about that Bag in a Fedora repository. An administrative and reporting interface will allow interaction with that content as well as access to services available to that preserved content. Restoration of content will initially be by initiating a restoration request and having scripts process content out of their preserved bags and placed back in the staging area in the format they were received for recovery by the original repository and staging tool.

## 6.4 Current System Architecture



## 6.5 Overview of Components

### 6.5.1 Staging Of Content for Ingest

Content to be ingested in APTTrust first needs to be copied from the originating repository or filesystem to a Duracloud space that will be used to process and ingest content from. Duracloud spaces will be provided for each repository option at each Partner Institution as needed to support content ingest. For for instance if UNIVERSITY EX has both a DSpace and a Fedora repository they will have a seperate Duracloud staging space for each named something like 'uex-dspace-staging' and 'uex-fedora-staging' respectively.

**Fedora & Cloudsync:** Partners with content originating in Fedora repositories will be asked to install [Fedora Cloudsync](#) in order to stage content in the repository specific Duracloud staging area.

**DSpace & Replication Task Suite:** Users of DSpace version 1.8.2 and later are asked to install the Replication Task Suite plugin to serialize and copy content as AIPs to their respective staging area.

**Filesystem & Curator's Workbench:** Partners with filesystem objects they wish ingested into APTrust should serialize their content into some standardized format using Curator's Workbench and copy the AIPs produced to the Corresponding Duracloud Staging area.

## 6.5.2 Content Processing

**Bagging and Fixity:** Scripts will process content from the Staging area into a defined BagIt structure for preservation in APTrust with fixity information stored in the Fedora Object with pointers to the bag.

**Gathering Metadata:** Relevant metadata digital objects will be parsed as the preservation bag is being created and updated as appropriate in the Fedora Object for that bag. This is experimental to see how much metadata we might be able to read and not a replacement for submission metadata from the original owner.

## 6.5.3 Administrative Interface

The administrative interface is the primary means by which partners will interact with and manage their content in APTrust and will include RESTful API management options as well. Using Hydra, it will allow the management of APTrust metadata for objects, allow administrators to view reports on the preservation activities around objects, and access services on content made available as part of APTrust.

## 6.5.4 Preservation Space

The primary cloud storage will use Duracloud to manage the fixity and duplication of content across geographically redundant regions. Digital objects will be stored as Bags serialized using the BagIt standard but the exact details need to be worked out and some investigation made into storage size limits and other areas before a final storage strawman can be created. The intent is to store content as bags with pointers to those bags from Fedora with fixity checks being performed by Duracloud and results recorded in Fedora.

## 6.5.5 Restoration Services

Since the intent is to store the serialized content received 'as is' in the data directory of the Bag, when content needs to be restored we will copy the contents of the bags data directory back to staging. This should allow partners to restore content to their repository using the same tool used to stage it in the first place.

## 6.5.6 Sending Content to DPN

Content marked as DPN bound will be pushed to DPN, which itself uses the Bagit structure to serialize content to a preservation format before replicating it across the DPN Federation of Nodes. DPN provides a variety of deep preservation and success services but in essence it is a Dark Archive. Content may be updated in DPN from APTrust but never deleted. Broader policies on how and when content should be updated in DPN is under development.

## 6.5.7 Earlier Architecture Proposals and Documentation

- [March 12th Meeting](#)

## 6.6 AP Trust Administrative Interface

### 6.6.1 `/*<![CDATA[*] div.rbtoc1486669573549 {padding: 0px;} div.rbtoc1486669573549 ul {list-style: disc;margin-left: 0px;} div.rbtoc1486669573549 li {margin-left: 0px;padding-left: 0px;} /*]]>*/`

- [Functionality](#)
  - [Content Ingest](#)
  - [Content Discovery](#)
  - [Content Modification](#)
  - [Reporting](#)
  - [Content Ingest Status](#)
  - [Export](#)
  - [Security](#)
- [Mock-ups](#)

The AP Trust administrative interface is a web application that exposes the functionality of the AP Trust repository to those administering content from the various partner institutions..

## 6.6.2 Functionality

### Content Ingest

1. Ingest will be initiated by [a client or external system](#) that may be specific to the ingest source (fedora vs. dspace vs. other) rather than having the technology dictate where the user goes for ingest.
2. In the initial version of the Admin interface, ingest process reporting will begin at the time content starts appearing in the DuraCloud staging area
  - a. Reporting should include time, user-id, and content-ids for each item moved to the staging space
  - b. Reporting should include time, user-id, and content-ids for each item which failed during upload to the staging space
  - c. Status should include...
  - d. Possibly at a later time, the tool that initiates and copies data to the cloud can report more information such as
    - i. a manifest for the entire ingest operation (so that data transfer progress can be computed and reported)
3. Show messages, error reports, and progress information for materials that are in the process of being ingested into the repository
  - a. Since the Discovery Service will only show content which has been ingested in the Aggr-Repo and indexed by Solr, the pre-ingest workflow state of any given item needs to be made available to the user
    - i. Workflow states are documented [here](#)

### Content Discovery

1. Allow user to search/browse based on any of the AP Trust [submission metadata](#) (beyond a certain point in the ingestion process)
  - a. i.e., Beyond the point where a valid submission metadata record has been processed.
  - b. The initial plan is to use a Solr Index (maintained by various components in the ingest process) as the sole index for discovery.

2. Allow user to answer the following questions
  - a. What's in there?
  - b. What did we put into APTrust during some date range?
  - c. Is an object with PID X in the repo?
  - d. What is everything that is in APTrust for my institution?
  - e. Why didn't this make it in?
  - f. How many things are in there?
  - g. What is the status of this ingest?
  - h. how much storage am I using?
  - i. how much and which content is DPN-bound?
  - j. how much is public/private/institution-only?
  - k. How much is the bill going to be this [timeframe]? **later**
3. Once content is discovered, a user should be able to view/download it (true or false?)

## Content Modification

1. Update not supported, **will be supported later**
2. Delete not supported, **will be supported later**

## Reporting

1. Various types of reporting will be ultimately available through this interface.
  - a. Audit - read: fixity checks
    - i. Initially a report on the periodic fixity checks will be available.
  - b. Storage, should be able to rely on existing DuraCloud storage reports
  - c. Provenance - an integrated log of what events/actions have taken place per item
2. Reports should be available for direct viewing in the admin console
3. Reports should be available for download
4. Allow for on-demand fixity checking. **later**
5. Support DPN end-user reporting functionalities?

## Content Ingest Status

1. Runtime ingest activity status will contain the following events
  - a. Ingest request started
  - b. Ingest request complete (with or without errors)
2. Each status event will show the following details
  - a. Date/Time of event
  - b. User-id of actor
  - c. Ingest identifier (user-defined name for ingest)
  - d. Number of packages included in the ingest

## Export

1. Exporting to public sources (fedora repositories) may be scheduled or triggered from the administrative interface. **later**

## Security

1. Provide some user authentication with simple permissions. (Shibboleth enabled eventually)
  - a. Users and roles will be established outside of this application
2. Need to support filtering content in UI based on logged-in user-id
3. Two roles need to be supported
  - a. ADMIN = institutional content administrator
  - b. ROOT = superuser with full access over content/messages/etc

### 6.6.3 Mock-ups

Page 152 of 190



University of Virginia - AP Trust Administrator

[http://aptrust.duraspace.org/discover?filter=dpn:true&terms:alpha or beta or gamma or delta](#)

AP Trust  
University of Virginia

Thomas Jefferson [Sign out](#)

[Dashboard](#) | [Discovery](#)

Filter results by

▼ Ingest Date Range  
from  /  to  /

▼ DPN Bound  
☒ Yes ☐ No

▶ Access  
▶ Health Check

Filtering on  
dpn: yes

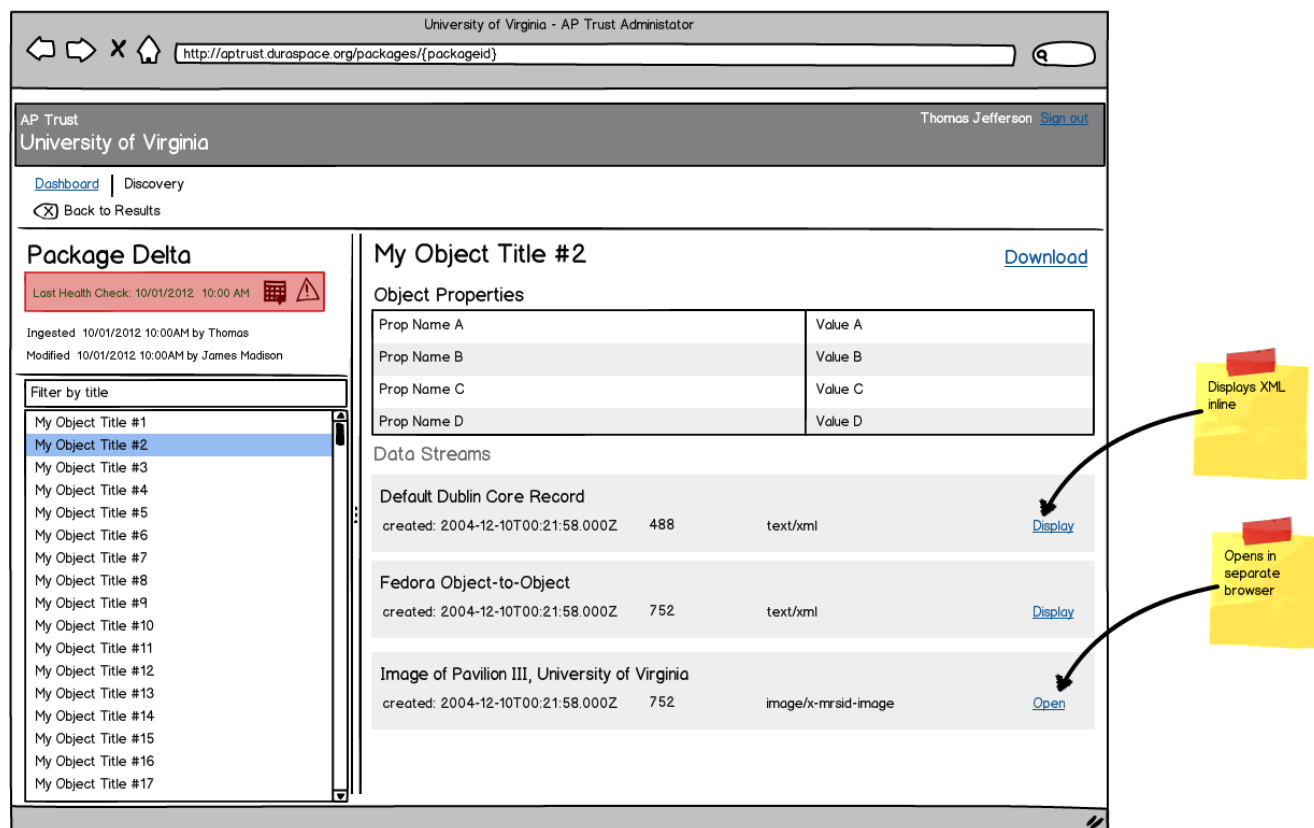
1-4 of 4

Package Alpha  
50 Objects | University of Virginia | Added 09/15/2012  
Last Health Check: 10/01/2012 10:00 AM

Package Beta  
50 Objects | University of Virginia | Added 09/15/2012  
Last Health Check: 10/01/2012 10:00 AM

Package Delta  
50 Objects | University of Virginia | Added 09/15/2012  
Last Health Check: 10/01/2012 10:00 AM

Package Gamma  
50 Objects | University of Virginia | Added 09/15/2012  
Last Health Check: 10/01/2012 10:00 AM



The screenshot shows the AP Trust Admin Interface for the University of Virginia. The top navigation bar includes 'Dashboard' and 'Discovery' links, along with a 'Back to Results' button. The main content area is divided into two sections: 'Package Delta' on the left and 'My Object Title #2' on the right. The 'Package Delta' section shows a 'Last Health Check' status and a list of objects. The 'My Object Title #2' section displays 'Object Properties' and 'Data Streams'. The 'Object Properties' table lists four properties: Prop Name A, Prop Name B, Prop Name C, and Prop Name D, each with a corresponding value. The 'Data Streams' section lists three streams: 'Default Dublin Core Record', 'Fedora Object-to-Object', and 'Image of Pavilion III, University of Virginia'. Each stream has a 'Display' or 'Open' link. Two yellow sticky notes with red tabs are placed on the right side of the interface. The first sticky note, labeled 'Displays XML inline', has an arrow pointing to the 'Display' link for the 'Default Dublin Core Record' stream. The second sticky note, labeled 'Opens in separate browser', has an arrow pointing to the 'Open' link for the 'Image of Pavilion III, University of Virginia' stream.

University of Virginia - AP Trust Administrator

http://aptrust.duraspace.org/packages/{packageid}

AP Trust University of Virginia

Thomas Jefferson [Sign out](#)

[Dashboard](#) | [Discovery](#)

[Back to Results](#)

**Package Delta**

Last Health Check: 10/01/2012 10:00 AM

Ingested 10/01/2012 10:00AM by Thomas

Modified 10/01/2012 10:00AM by James Madison

Filter by title

- My Object Title #1
- My Object Title #2**
- My Object Title #3
- My Object Title #4
- My Object Title #5
- My Object Title #6
- My Object Title #7
- My Object Title #8
- My Object Title #9
- My Object Title #10
- My Object Title #11
- My Object Title #12
- My Object Title #13
- My Object Title #14
- My Object Title #15
- My Object Title #16
- My Object Title #17

**My Object Title #2** [Download](#)

**Object Properties**

Prop Name A	Value A
Prop Name B	Value B
Prop Name C	Value C
Prop Name D	Value D

**Data Streams**

<b>Default Dublin Core Record</b>			
created: 2004-12-10T00:21:58.000Z	488	text/xml	<a href="#">Display</a>
<b>Fedora Object-to-Object</b>			
created: 2004-12-10T00:21:58.000Z	752	text/xml	<a href="#">Display</a>
<b>Image of Pavilion III, University of Virginia</b>			
created: 2004-12-10T00:21:58.000Z	752	image/x-mrsid-image	<a href="#">Open</a>

Displays XML inline

Opens in separate browser

## 6.6.4 Backend Needs for Admin Interface

Given the design of the features and appearance specified in the wireframes for the Admin Interface, the back-end support can be determined.

### Solr Index

#### Field List

- record\_type: value of "operation", "package" or "pid"
- institution\_id: values corresponding to the institution IDs
- dpn\_bound: true or false
- acces\_control\_policy: one of the enumerated policy values
- failed\_health\_check: true or not present
- operation\_status: one of the enumerated operations status values (TBD)
- date\_sort: a field that allows sorting by date
- ingest\_date: date/time ingest was completed
- keyword: any text we want keyword-searchable
- contained\_pid: a pid contained in a given package

- `modified_date`: the date/time a package was last modified

## Questions

1. For each `record_type` above, which fields will be available?

interface need	solr fields used	Direct Solr Query	Rest API call
number of packages ingested per institute	<code>record_type</code> , <code>institution_id</code>	<b>+record_type:package</b> <b>+institution_id:[id]</b> (and get the <code>/response/result/@numFound</code> )	GET <code>/rest/[id]/summary</code>
number of objects ingested per institute	<code>record_type</code> , <code>institution_id</code>	<b>+record_type:object</b> <b>+institution_id:[id]</b> (and get the <code>/response/result/@numFound</code> )	GET <code>/rest/[id]/summary</code>
number of GB used per institute	NOT FROM SOLR		GET <code>/rest/[id]/summary</code>
number of DPN-bound packages per institute	<code>record_type</code> , <code>institution_id</code> , <code>dpn_bound</code>	<b>+record_type:package</b> <b>+institution_id:[id]</b> <b>+dpn_bound:true</b> (and get the <code>/response/result/@numFound</code> )	GET <code>/rest/[id]/summary</code>
number of public packages per institute	<code>record_type</code> , <code>institution_id</code> , <code>access_control_policy</code>	<b>+record_type:package</b> <b>+institution_id:[id]</b> <b>+access_control_policy:world</b> (and get the <code>/response/result/@numFound</code> )	GET <code>/rest/[id]/summary</code>
number of private packages per institute	<code>record_type</code> , <code>institution_id</code> , <code>access_control_policy</code>	<b>+record_type:package</b> <b>+institution_id:[id]</b> <b>+access_control_policy:private</b> (and get the <code>/response/result/@numFound</code> )	GET <code>/rest/[id]/summary</code>
number of institution only packages per institute	<code>record_type</code> , <code>institution_id</code> , <code>access_control_policy</code>	<b>+record_type:package</b> <b>+institution_id:[id]</b> <b>+access_control_policy:institution</b> (and get the <code>/response/result/@numFound</code> )	GET <code>/rest/[id]/summary</code>
number of packages with failed health checks	<code>record_type</code> , <code>institution_id</code> , <code>failed_health_check</code>	<b>+record_type:package</b> <b>+institution_id:[id]</b> <b>+failed_health_check:true</b> (and get the <code>/response/result/@numFound</code> )	GET <code>/rest/[id]/summary</code>

interface need	solr fields used	Direct Solr Query	Rest API call
current ingest operations (id, name, start time, starting user, progress)	record_type, operation_status, date_sort	<b>+record_type:ingest +institution_id:[id] -operation_status:complete</b>	
failed ingest operations (id, name, start time, starting user, failure notice)	record_type, operation_status, date_sort	<b>+record_type:ingest +institution_id:[id] +operation_status:failed</b>	
completed ingest operations (id, name, start time, package count, size)	record_type, operation_status, date_sort	<b>+record_type:ingest +institution_id:[id] -operation_status:ingest_complete</b>	
user query with facets	record_type, dpn_bound, ingest_date, failed_health_check, access_control_policy, keyword		
package display and filtering	record_type, contained_pid, ingest_date, modified_date POSSIBLY QUERY FEDORA DIRECTLY		

## Questions

1. Will the number of objects per package be stored in solr? or will method #2 require parsing the manifest?
2. For method #13, is the health\_check info **and link** available with each "package" record?

## 6.7 APTrust Data Storage Options



An executive decision was made by those present at the 2012-09-24 face-to-face meeting that fedora would not have large managed datastreams, but instead that content would live in the preservation cloud storage and be "Externally Referenced" datastreams. Foxml would also exist (in backup) on the preservation storage.

This page is a reference to the arguments for and against a historical decision that has since been made.



### Warning

The workflow or architecture described here are potentially out of date. This page serves only to document the decision-making process.

Information about the current AP Trust architecture can be found [here](#).

### 6.7.1 Storage Goals

With a given that the data will be managed by the Fedora aggregation repository, there are several possible locations and management scenarios for content in AP Trust. Each poses slightly different development requirements and has implications on ingest performance and monetary expense. The following project goals, requirements or suspected use cases can be used to contrast the various approaches.

1. content that is fully ingested into AP Trust will need to be "preserved"
2. content that is fully ingested into AP Trust will need to be subject to periodic and automated checksums to audit file fixity
3. content that has fully or partially ingested into AP Trust must be able to be retrieved
4. AP Trust must be able to handle large data files
5. AP Trust must be able to handle a large volume of parallel ingests

	Externally Referenced Approach	Fedora Managed Approach
--	--------------------------------	-------------------------

	Externally Referenced Approach	Fedora Managed Approach
description	Large content is not managed by fedora, but instead is an external reference URL ("E") pointing to DuraCloud.	All content is managed directly by fedora, but a custom Akubra storage plugin is used to put datastreams in DuraCloud storage while foxml is in the local fedora storage.
data flow	Pre-ingest routine prepares FOXML and datastreams CloudSync moves FOXML and datastreams into DuraCloud (initiated by user) CloudSync copies/ingests augmented FOXML into Fedora while leaving large datastreams in the cloud (schedule and asynchronous to maximize performance)	Pre-ingest routine prepares FOXML and datastreams CloudSync moves FOXML and datastreams into Fedora
preservation (1)	If the storage for the fedora aggregation repository isn't replicated geographically with durability guarantees, the FOXML will have to be retained in the DuraCloud storage (and possibly synchronized with fedora)	Does this work if our storage is replicated between Amazon and San Diego?
automated checksum reports (2)		The DuraCloud services will have to operate against the same data that fedora is managing, and therefore be limited to read-only operations to take advantage of the existing audit functionality.
retrievability (3)		
large files (4)	Work will have to be done to allow CloudSync to transfer large files.	Work will have to be done to allow CloudSync to transfer large files.
large volume (5)		It's possible (but not guaranteed) that fedora ingest will be a bottleneck.

## 6.8 AP Trust Participant Setup

For every participant in AP Trust:

1. A [brief identifier](#) should be assigned

2. A staging and production space should be created in duracloud
  - a. the production space should be identified with just the institution code (ie, uva)
    - i. the property "type" should be set to "preservation\_store"
    - ii. the property "institution\_display\_name" should be set to the human readable and preferred name of the institution (ie, "University of Virginia Libraries")
  - b. the staging space should be identified with the institution code followed by "staging" (ie, uvastaging)
3. The institutional administrator for the institution should be asked to create a DuraCloud account, which should be granted permission on the staging space

## 6.9 Architecture Overview (March 12)

---

### 6.9.1 Thoughts evolving from 21 March meeting at UVa Library (representatives of Duraspace and UVa present):

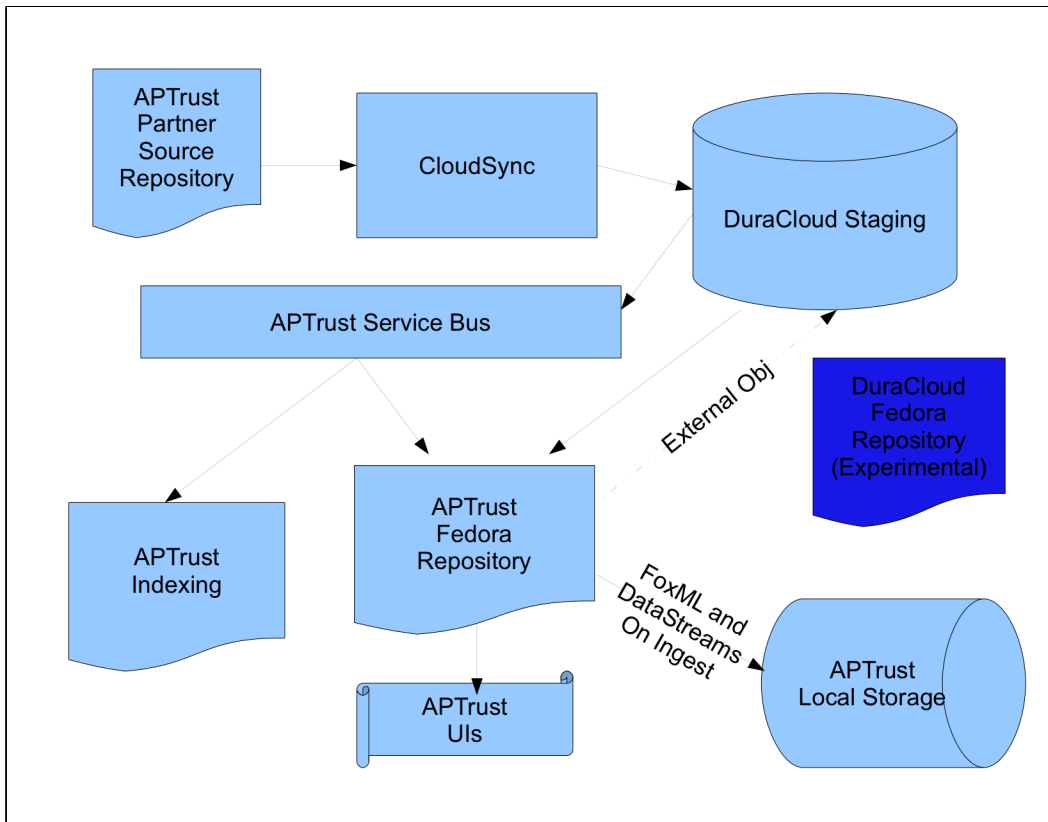
We have available [a few pictorial aids](#) showing the direction of thought.

Here are some questions they might raise:

- If we include CloudSync as the staging area / transport management component between repositories, how difficult will it be to support non-Fedora repositories?
- If we contemplate operating the APTTrust Fedora repository using both "as-good-as-local" storage and DuraCloud storage, will we require a rule-execution component in Fedora's storage subsystems to make that hybrid work? Is there some other, less costly way to make that hybridization?
- If we center the architecture on a service bus, should we expect to expose the APTTrust APIs and notifications directly from that bus? Should we expect to house the executing instance of DuraCloud in that bus? Should we expect to house supporting index services in that bus?
- Will the APTTrust APIs represent a superset of the Fedora APIs, or something very different?
- To what extent could we reuse (parts of) CloudSync's administrative interfaces? To what extent do CloudSync's notions overlap with APTTrust's, at the level of administrative use cases?
- If contributing institutions contribute out of non-Fedora repositories, can we make workflows in the service bus responsible for translating into the Fedora object model?

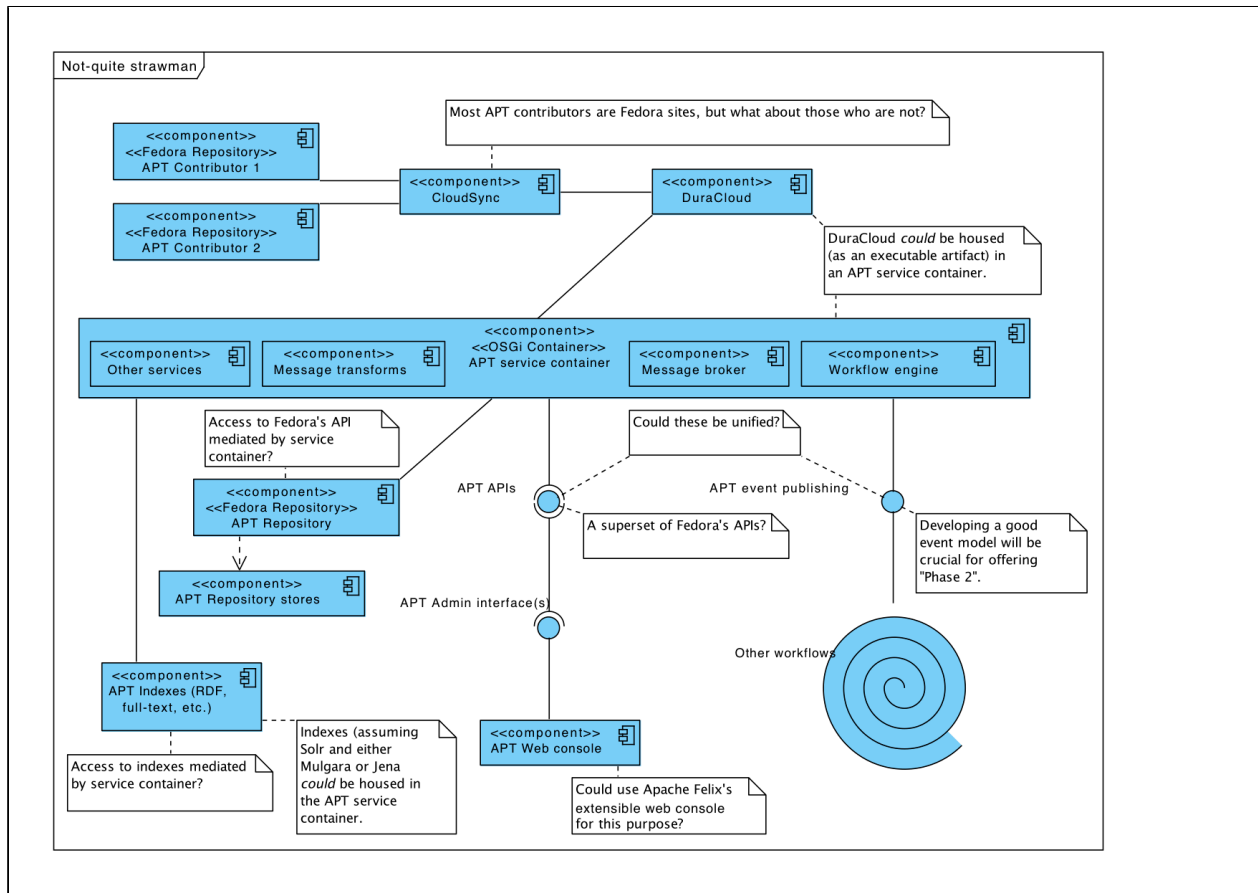
### 6.9.2 APT Architecture Sketches

1) Brad McLean's first rough sketch:



2) Adam Soroka's response:





## 6.10 Export Processing Service

The AP Trust "Export Processing Service" is a service running in DuraCloud that is responsible for preparing and possibly initiating the export of content from the AP Trust aggregation repository.

## 6.11 Ingest Client

To initiate ingest of content, especially content that isn't available on public servers, the user must run some client. Our goal is to leverage existing tools as much as possible, but some development work is expected. We do not wish to branch any existing tools, but if need be, modify them to allow for the types of general interactions necessary

### 6.11.1 Ingest Client Data Flow (simple)

1. Information about what materials are to be ingested including the AP Trust submission metadata is provided to the Ingest Client
2. The Ingest client prepares a manifest and puts it in the AP Trust staging area

3. The ingest client invokes another repository specific tool or service to transfer the materials from the client repository to the AP Trust staging area.



The tool used to transfer the content (and possibly the manifest) would likely be a copy of Fedora CloudSync for fedora content, the [ReplicationTaskSuite](#) for DSpace content and the possibly [Duracloud Sync Tool](#) or rsync for filesystem for getting content to the DuraCloud storage.

## 6.11.2 Ingest Client Data Flow (sophisticated)

1. Information about what materials are to be ingested including the AP Trust submission metadata is provided to the Ingest Client
2. The Ingest Client uses the provided information and queries the AP Trust system to determine whether there are any conflicts with concurrent operations and take appropriate countermeasures (waits its turn, halts ingest with a helpful message, etc.)
3. The Ingest Client uses the provided information and queries the AP Trust system to determine if any optimizations can be made (like not transferring files that haven't been modified, updating metadata in the case of metadata-only changes, etc.)
4. The Ingest Client prepares a manifest and puts in in the AP Trust staging area
5. The Ingest Client invokes another tool to transfer the materials to the AP Trust staging area

## 6.11.3 Functional Requirements

An "ingest client" may involve any sort of interaction (or lack of interaction) with a user or content but must result in:

1. An AP Trust ingest Manifest is provided to the AP Trust system
2. All materials referenced as added or updated by the above Manifest in that same AP Trust staging cloud storage space.

In order for useful display of ingest progress in the Admin Interface, the Manifest should appear first, or at least very early on since the ingest operation will not be visible until the manifest has been processed.

## 6.11.4

## 6.11.5 Ingest Manifest

In order to support the tracking of progress, a requirement for any ingest operation is the creation and transmission of an "ingest manifest". This manifest is created and transmitted by the Ingest client and though possibly visible in the staging space of the AP Trust cloud storage, it can be considered an internal format. This metadata is created by ingest clients and consumed by the Ingest Processing Service.

The successful transmission and processing of the manifest from an ingest client to the ingest processing service serves to:

- provide enough information so that the Ingest Client can prevent concurrent ingest operations from interfering with each other (possibly prevent all concurrent ingest operations in the simplest case)
- provide information that is passed along to the admin interface to allow information on a current ingest operation (progress, size, etc.)
- indicate the status of an ingest operation (that it's begun, but isn't complete, since presumably the [IPS](#) would move or update the manifest when the transfer was complete)
- trigger the deletion of materials
- trigger metadata changes to materials

Because XML processing is relatively easy in Java, Fedora is well suited to store XML data and there exist languages to document new XML structures (XSD, schema), it seems that the manifest should probably be an XML representation of the needed information.

Currently, this manifest could contain:

- a human readable name for the ingest operation (suggested by Greg J)
- the user who initiated the ingest
- the time the ingest was initiated
- the list of packages that are part of the ingest with:
  - list of files that make up the package
  - sizes (if available)

**sample-manifest.xml**

```
<?xml version="1.0" encoding="UTF-8"?>
<manifest xmlns:fedora-aptrust="http://www.aptrust.org/relationships#" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <!--
    This AP Trust manifest represents content being ingested into AP Trust. It may
    or may not refer to "packages" explicitly defined by the user prior to ingest, but
    once processed will result in one or more "packages" in the AP Trust system.

    Currently a "package" is a unit at which AP Trust administrative metadata is
    applied. For the most simple manifest files, a single package is implied. For
    more complex manifests, various strategies may be employed.
```

This manifest is created by the Ingest Client. This manifest and the files it references are the sole mechanism **for** conveying information to the Ingest Processing Service that processes content to be ingested into the AP Trust aggregation repository.

In other words, when a valid manifest file and referenced data files are found on the AP Trust staging space **for** an institution, they will result in the content being ingested (any any supplementary **"packages"** created).

-->

<description>

<!-- Any name supplied by the initiating user, or some **default** in the **case** where such an interaction is too onerous. -->

<name>Initial Ingest</name>

<!-- We might be able to get the username automatically from the authentication against the cloud storage space, in which **case this** is either irrelevant or less trusted. -->

<suppliedUsername>md5wz</suppliedUsername>

<!-- The time the ingest was initiated may be different from the time first file arrives on the server. This self-reported time may be most useful in the user **interface**, but could be incorrect -->

<aptrust>yyyy-MM-dd HH:mm:ss.sss</aptrust>

</description>

<packages>

<!--

For object in fedora where packaging (and AP Trust metadata) will not be maintained in the institution's local repository.

-->

<package>

<!-- An object will be created in the AP Trust aggregation repository to encapsulate **this** metadata and link to the referenced objects. This **package** will be assigned a PID in the AP Trust repository when it's created.

-->

<metadata rdf:about="**test:1**">

<fedora-aptrust:hasAccessControlPolicy>restricted</fedora-aptrust:hasAccessControlPolicy>

<fedora-aptrust:isDPNBound>**false**</fedora-aptrust:isDPNBound>

<fedora-aptrust:title>Example title</fedora-aptrust:title>

</metadata>

<digitalObjects>

<digitalObject type="**FEDORA**" version="**3.5**">

<size>-1</size>

<pid>uva-lib:**602146**</pid>

</digitalObject>

<digitalObject type="**FEDORA**" version="**3.5**">

<size>-1</size>

<pid>uva-lib:**602179**</pid>

</digitalObject>

<digitalObject type="**FEDORA**" version="**3.5**">

<size>-1</size>

```
<pid>uva-lib:602193</pid>
</digitalObject>
</digitalObjects>
</package>

<!-- An example DSpace Package -->
<package>
  <digitalObjects>
    <digitalObject type="DSpace" version="1.8.2">
      <size>679400</size>
      <pid>hdl:xxxxx/y</pid>
    </digitalObject>
  </digitalObjects>
  <metadata rdf:about="hdl:xxxxx/y">
    <rdf:hasAccessControlPolicy>restricted</rdf:hasAccessControlPolicy>
    <rdf:isDPNBound>false</rdf:isDPNBound>
    <rdf:title>Title of DSpace package</rdf:title>
  </metadata>
</package>
</packages>

</manifest>
```

## 6.12 Ingest Processing Service

The AP Trust "Ingest Processing Service" is a proposed service to be run in DuraCloud that detects uploads to the institutional ingest staging space and performs any remaining work to ingest the content into the AP Trust fedora aggregation repository.

### 6.12.1 Functionality

- respond automatically to the presence of [manifest files](#) to:
  - summarize "in-progress" ingest or update operations in the index (Solr) to support the [Admin interface](#) and processing of forthcoming files
    - creates a record in Solr for the current ingest operation that can be queried by the Admin Interface to get the information about "Active Tasks"
    - registers the manifest, it's content and whatever else is needed by the Ingest Processing Service to support quick identification and processing of files referenced in the manifest

- **respond automatically to the presence of recognized (ie, in a current manifest) files**
  - summarize ingest or update operations to support the admin interface
    - update the record in Solr for the current ingest operation to support the progress bar in the Admin Interface
  - validate and process files
    - ensure that provided files represent complete and sufficiently described submission packages
    - for non-fedora content, create a fedora object to wrap the content
    - ensure that the submission of the given packages is allowable given data already in the repository
      - disallow overwriting of DPN-bound content
      - disallow packages that reference resources that are already part of another package
      - disallow packages whose identifiers may conflict with existing packages
  - move complete packages (which originated as fedora objects or were wrapped in fedora objects) into the aggregation space and ingest them into the aggregation Fedora repository
    - if errors occur putting the objects into fedora, roll back the entire package operation
  - remove processed files and manifests from the staging space once processing and ingest has been completed
- **respond automatically to the presence of unrecognized files**
  - behavior is not yet defined but could range from ignoring the content to providing information to some reporting mechanism

## 6.12.2 REST API

To query information needed by other components in the AP Trust system or to manually trigger some operation (if manual operations are available) a simple REST API will be available. The details will emerge as the application is implemented.

## 6.12.3 Ingest Workflow State

For simplicity and to avoid concurrency problems between clients, the administrative tool and the IPS itself, ingest workflow state may be mediated by this service.

The following is a rough idea of the states an ingest could be in:

State	Implications	Location of input data	Location of output data	Where this state is stored

State	Implications	Location of input data	Location of output data	Where this state is stored
Transfer to staging in progress	No conflicting transfers may be initiated.	partner institutions		an ingest manifest
Transfer to staging failed	User will want to know about this.			?
Transfer to staging complete	IPS will be triggered automatically when files arrive and will process pieces based on information in the manifest, or do nothing if no manifest has arrived.		institutional cloud staging space	?
IPS processing in progress		institutional cloud staging space		?
IPS processing failed	User, and APTrust admin will want to know about this.			?
IPS processing complete	CloudSync will be triggered		institutional cloud processed space	?
Fedora Ingest in progress		institutional cloud processed space		?
Fedora Ingest failed	User will want to know, and we might need to roll back the data change in the "institutional cloud processed space"			?
Fedora Ingest complete	Index update will be triggered automatically.			?

State	Implications	Location of input data	Location of output data	Where this state is stored
Indexing complete	This implies complete ingest of the object.			Solr

There may need to be roll-back states in the above, for instance if an object fails in the last state, we'll need to synchronize the institutional cloud processed space.

## Maintenance of State information

It's not clear where the state should be maintained, but there are the following requirements:

- It must be persistent, such that a restart of any or all of the services can be tolerated to the highest degree possible.
- It must support synchronous updates to the extent that it must be relied on by the various components. This may or may not include locking.
- It must be ultimately preserved as part of the provenance record for the object in AP Trust
- It must be easily queried for presentation to users in the admin interface.

### Possible stores

- fedora
- solr
- some sort of database
- files in the cloud
- some combination of the above

## 6.13 Submission workflow and tools

In order to get content into the AP Trust aggregation repository, some work must be done at each institution. To foster easy and efficient participation, part of the work done for the phase 1 implementation of AP Trust is to develop tools and workflows that are suitable for the institutions involved. The major milestones in that development effort include:

1. develop a model for the AP Trust content in the repository
2. develop a procedure to get sample content from a UVA fedora repository into the test AP Trust repository
3. develop a procedure to get sample content from UVA that is outside of a fedora repository into the test AP Trust repository
4. develop routines go get content from all partner institutions that are willing to submit sample data into the test AP Trust repository



In order to better understand the variety of ingest scenarios for step 4 above, [a survey](#) will be sent to all participants.

## 7 Best Practices

---

- [Coding Styleguide](#)
- [Pivotal Tracker](#)
- [Service Setup and Configuration](#)
  - [Demo Configuration and Setup](#)
- [Version Control](#)

### 7.1 Coding Styleguide

---

#### 7.1.1 Overview

This document contains information about standard practices and agreed upon Best Practices used on the project throughout the various codbases.

Our standard best practice is currently to code to Java 6 using the latest version of Maven for project management.

- [Overview](#)
- [Ruby](#)
  - [Documentation](#)
  - [Hashes](#)
- [Python](#)
  - [Documentation](#)
- [Java](#)
  - [Package Conventions](#)
  - [Identifying Software Versions](#)
  - [Spaces or Tabs](#)

#### 7.1.2 Ruby

Unless otherwise specified below, please follow the recommendations on the [GitHub Ruby Style Guide](#) when coding in Ruby.

#### Documentation

Until we decide otherwise use the rDoc spec for docstring comments.

## Hashes

Use the Hash Literal method introduced in Ruby 1.9.

```
# good
robot = {
  name: "Optimus Prime",
  type: "Autobot",
}

# bad
robot = {
  name: "megatron",
  type: "decepticon",
}
```

## 7.1.3 Python

Unless otherwise specified below, please follow the standard [PEP-8](#) recommendations.

## Documentation

Please use [Restructured Text](#) format in docstring comments and used the [Sphinx documentation library](#) for overall documentation of specific libraries and codebases.

## 7.1.4 Java

### Package Conventions

Java packages for APTTrust should follow the naming convention ``org.aptrust.<package>`` where `<package>` is named or extended as appropriate.

### Identifying Software Versions

For Java applications we are using Maven for project management

```
<major version>.<minor version>.<incremental version>--<qualifier>
```

In the case of Maven projects `<qualifier>` should use the standard 'SNAPSHOT' qualifier for code in development.

### Spaces or Tabs

All code pages should use spaces. Tabs should never be used.

## 7.2 Pivotal Tracker

---

### 7.2.1 Overview

For general project management, APTrust will be using the agile project management online application [Pivotal Tracker](#).

- [Overview](#)
- [Using Pivotal Tracker](#)
  - [Tutorials](#)
  - [Signing Into Pivotal Tracker](#)
- [User Stories](#)
- [Story Points & Estimation](#)
  - [Ideal Sample User Stories](#)
- [GitHub Post Commit Hooks](#)

### 7.2.2 Using Pivotal Tracker

We will use Pivotal Tracker for both the central APTrust application, related DPN code for APTrust and any additional projects or collaborations that may come along. Epic Stories will form the primary goals of the we're trying to achieve from which we will create User Stories and estimate them with Story Points using a Fibonacci scale based on their complexity. Iterations are 2 weeks long

### Tutorials

Quiet a bit of information exists already to help people ease into Pivotal Tracker and so instead of recreating it I will link that here for your reference. In addition to the resources below feel free to check out the [Pivotal Tracker YouTube](#) channel as well as the [Pivotal Tracker Help](#) pages.

- I ***highly recommend*** you view the 8 Minute [Introduction to Pivotal Tracker Concepts](#) on YouTube.
- There is also an excellent 2 minute video [Introduction to Epic User Stories](#) that I recommend.
- There is another 8 minute video you could optionally watch as well that covers [Advanced Pivotal Tracker](#) use.

### Signing Into Pivotal Tracker

I recommend you use your Google account to sign into Pivotal Tracker for consistency (see diagram below). To do that just click on the google apps icon at the bottom of the login screen and consistently use that so you wont have to create separate accounts. If you don't have one or don't want to use it you can always create a local Pivotal Tracker account as you see fit.

*Link to sign in using Google account.*



## 7.2.3 User Stories

A good User Story should form a narrative with a very brief description of "[**WHO**] is doing [**WHAT**] and [**WHY**]" in an application so Engineers can break this down into individual tasks necessary to complete the User Story.

The **WHO** of a user story can really be anything and isn't just limited to a person. It can be an application or script, a particular kind of user (like an authenticated one) or something very abstract like a website visitor. The who of a user story should help keep the context of an operation clear and imply the needed roles and security needed for that feature.

The **WHAT** of a user story should be rather clear and ultimately testable in the end so that someone unfamiliar with the development of this user story can confirm if this request has actually been satisfied and the application or feature works as requested.

The **WHY** of a user story should give some short perspective on how this feature fits in with the overall user experience or functionality of the application and why it is important.

Rather than recreate the wheel, descriptions of good user stories exist abound on the web already. Google for it a bit or check out:

- [Agile Software Development: User Stories](#)
- [Agile Project Planning: Writing Good User Stories](#)
- [Advantages of a User Story Template](#) (with nice bit on formatting the story)

### *Some Tips*

- Try to stick to granular, deliverable and testable features requests.
- If you can't say it in a single sentence, what you're trying to do is probably too complex and needs to be broken down.

- User Stories should form the basis of features to be bundled in releases of your product.
- If you don't understand what you just wrote, odds are, neither does anyone else.
- The WHY of a User Story is often the most neglected but ultimately the most important part of a feature, always know and be able to state why a story is needed.

### Example of Creating User Stories in a Project

As an APTrust Institutional Admin I can easily a list of my content in APTrust to identify items that are DPN bound so determine if it has been successfully stored in that Archive.

As an APTrust Institutional Admin I can click on a link for any particular item in APTrust to queue that item for placement in the staging area so I can restore original content lost or corrupted in my repository.

## 7.2.4 Story Points & Estimation

User Stories are estimated for their general complexity relative to other User Stories in a project and assigned a Story Point value based on that estimation to help communicate how large a story is and to help estimate how long it will generally take to complete that feature. Story points are estimated *collaboratively* by the development group for relative complexity against 'ideal Sample User Stories' other user stories in the project for consistency.

More to come

### Ideal Sample User Stories

more to come

## 7.2.5 GitHub Post Commit Hooks

In most cases we'll have activated GitHub Service Hooks for PivotalTracker in each project. If active you should be able to add the ID number of the story or issue you're fixing or working on in the commit message and it will automatically update the appropriate ticket once it is pushed to GitHub.

In your commit message just start it with "[#<ticket id>]" and it should update that ticket. For instance if you have an issue ID number 47853987 then just start the commit message with that and add whatever other text you want like so

```
git commit -m "[fixes #47853987] Skynet is now self aware"
```

And the appropriate ticket should be updated. Multiple tickets can be added to the same commit message by putting spaces between the message numbers.

See article on [SCM Post-Commit Message Syntax](#) for more options.

## 7.3 Service Setup and Configuration

### 7.3.1 Setting Up a Server and Secondary Server

All of the software needed for the April 29 demonstration exists on two servers: (1) APTrust demonstration server and (2) A DuraCloud instance (Durastore) server.

The APTrust demonstration server contains all the software currently in development. Not all this software is needed for demonstration of the basic ingest processor and GUI. The important components that are not needed are RabbitMQ and CloudSync. Technically Fedora is not needed but it takes work to deconfigure it and it does not cause problems.

Host: [aptrust.duracloud.org](http://aptrust.duracloud.org)

Software:

All the software except RabbitMQ is located in `/opt/fcrepo` and deployed in the same Apache Tomcat instance that is installed by Fedora. The environment variables are set from the `/host/fcrepo` user and found in `/home/fcrepo/.bash_profile`. Since it is important to avoid permissions problems, the developer should always log into the `fcrepo` user to cycle Tomcat. Currently, that may be accomplished by using `ssh` to access the [ubuntu@aptrust.duracloud.org](mailto:ubuntu@aptrust.duracloud.org) EC2 server using a shared keypair. Then `sudo -i` and `su fcrepo`. The common Tomcat can also be cycled via the script found in `/etc/init.d/fcrepo`. This script automatically sets the correct context so it can be run directly from any user.

Note: Propose adding the Tomcat manager application to the default AMI, Could use common account /password, individual accounts/password, or integration with ID service.

Note: Consider using Maven as a deployment mechanism. Note mutually exclusive with Tomcat manager.

Software	Needed	Startup	Shutdown	Notes
<aptrust> /aptrust-admin	Yes	Automatically with Tomcat	Automatically with Tomcat	Requires a configuration file or remote configuration to function
<aptrust> /cloudsync	No	Automatically with Tomcat	Should be individually shutdown	Only causes problems when two server instances are running and pointing to the same DuraCloud space<<
<aptrust> /examples	No	Automatically with Tomcat	Automatically with Tomcat	No action needed

Software	Needed	Startup	Shutdown	Notes
<aptrust> /fedora	Not Strictly	Automatically with Tomcat	Automatically with Tomcat	Requires deconfiguration to remove but no harm if left alone. Minimally must have <fedoraServerHost> property set to public hostname or IP address in /opt/fcrepo/server/config/fedora.fcfg for each EC2 instance.
<aptrust> /fedora-demo	No	Automatically with Tomcat	Automatically with Tomcat	Used to test Fedora, unneeded but causes no harm if left alone
<aptrust> /fop	No	Automatically with Tomcat	Automatically with Tomcat	A Fedora service, unneeded but causes no harm if left alone
<aptrust> /host-manager	No	Automatically with Tomcat	Automatically with Tomcat	A Tomcat service, not currently installed but a significant convenience during development
<aptrust> /imagemanip	No	Automatically with Tomcat	Automatically with Tomcat	A Fedora service, unneeded but causes no harm if left alone
<aptrust> /manager	No	Automatically with Tomcat	Automatically with Tomcat	A Tomcat service, not currently installed but a significant convenience during development
<aptrust> /saxon	No	Automatically with Tomcat	Automatically with Tomcat	A Fedora service, unneeded but causes no harm if left alone
<aptrust> /solr	Yes	Automatically with Tomcat	Automatically with Tomcat	
<aptrust> RabbitMQ	No	sudo rabbitmqctl start	sudo rabbitmqctl stop	Standalone. No evidence it is started automatically. Needed for DPN integration (and demos) but not for APTTrust. Could cause problems if started on two servers without proper configuration.
<duracloud>	Yes	By DuraSpace Staff	By DuraSpace Staff	Deployed and managed via DuraSpace administrative software. Can be shared between servers but care needs to be taken with named space or object name collision.



## 7.4 Overview

---

By default we use the [Git](#) version control system hosted under the [APTrust GitHub Organization](#). Because of highly dependent nature of libraries and code being used in the project we will generally manage code under a single Git repository with a Maven structure, with necessary POM.xml files to build code as needed. In cases where it makes sense like prototype or with completely disconnected codebases we may pursue separate repositories but for now we will use a single primary repository to be noted and linked here after it is created.

Currently all GitHub repositories are publicly accessible so please keep this in mind while coding.

### 7.4.1 Branching and Tagging Strategy

Overall we recommend a standard Master/Develop/Release/Tag strategy as explained in [This Post at Nive.com](#). In a nutshell:

- Master: Should contain the most recent stable codebase. Tags should be created from code merged back to Master from Develop to
- Develop: Should be for stable development code and generally use for integration and merging from release branches. Develop should see frequent merges coming from release branches as needed.
- Release: Generally a non-stable local release branch.
- Tags: Serve as the formal version-ed releases of a codebase.

[GitFlow](#) is a great Git extension that will essentially help manage the branching and tagging strategy for you. I've gotten a lot of use out of it and would recommend it.

## **8 Project Information**

---

### **8.1 Purpose**

---

The Academic Preservation Trust is a new organization and implementation to support the longterm preservation of scholarly work. APTTrust member organizations are funding APT1.0 which encompasses the formation of the new organization and the initial implementation of a preservation repository which will function as one part of a national federated preservation network.

### **8.2 APT 1.0 Guiding Documents**

---

#### **8.2.1 General Assumptions**

#### **8.2.2 APT 1.0 Implementation Project Charter**

#### **8.2.3 APT 1.0 Implementation Project Requirements**

#### **8.2.4 Communication Plan**

#### **8.2.5 Navigate space**

### **8.3 APT1.0 in Detail**

---

#### **8.3.1 APT 1.0 Implementation Project Plan**

## 8.3.2 APT 1.0 Implementation Timelines

## 8.3.3 APT 1.0 Roadmap and Outstanding Work

## 8.3.4 APT 1.0 Implementation Project Lessons Learned

## 8.4 APT1.0 Outstanding Work

This was adapted from a google doc spreadsheet that was shared. Maintaining it here may allow for greater centralization of project documentation.

Deliverable	Description	Who's responsible	Due date	Comments
Test content		UVa/liaison		Include Fedora and non-Fedora content; test data from UVa and other institutions.
	Define metadata	Bradley	6/30/2012	Already done? Check with Greg R
	Prepare UVa Fedora content	Andrew /Bradley	7/15/2012	Identify use cases as content is prepared (10)
	Prepare UVa non-Fedora content	Andrew /Bradley	7/30/2012	Identify use cases as content is prepared (10)
	Prepare UNC Fedora content	Andrew /Bradley /Greg J	8/15/2012	tentative date - idea is to move progressively through different sources & types of data
	Prepare NCState non-Fedora content	Andrew /Bradley /Greg R	8/30/2012	tentative date - idea is to move progressively through different sources & types of data
Ingest workflow		UVa/Chris		Will move in parallel with test data availability

	Flesh out ingest steps for UVa	Mike & Bradley	7/15/2012	Need use cases (4-7)
	Set up ingest environment	Mike	7/15/2012	
	Run UVa content through model	Andrew /Mike	7/30/2012	
	Determine adjustments to cloudsync	Mike/Chris	7/30/2012	Implemented when setting up cloudsync (24)
	Document workflow	Mike & Bradley		ongoing
	Push content into Cloud from UVA	UVa/Chris	7/30/2012	start date, could take few days or few weeks
DuraCloud OS		DuraSpace		
	Select cloud storage & server host	done	7/1/2012	Amazon for now, dual upload to SDSC once real content coming in
	Establish cloud	Robin /Michele /Bill	7/30/2012	Spin up cloud
	Investigate local compute & storage	Tim & Martha	8/30/2012	Can we provide a local cloud infrastructure for APTTrust? Will this make business sense? Technological sense?
	Move content into cloud	Mike/Bill	8/30/2012	Mike will signal readiness to begin data transfer when ingest activity permits
Cloudsync		DuraSpace		
	Install in cloud	Chris	8/15/2012	
	Implement modifications	Chris	8/30/2012	modifications identified during ingest process (13)

Fedora3.5 repository	Establish repository	DuraSpace	mid-August	
	Set up in cloud	Chris		
	Sync Foxml/xml into repository	Chris/Mike		
DEMO model w /out admin interface			9/30/2012	Without administrative interface (at ARL meeting in Oct)
Admin interface		All		
	Define services	Bradley, Donna, Greg R	7/15/2012	ingest, retrieve, modify, audit, delete, others? Needed for ingest workflow
	UI design	?????	September	
	Implement Shibboleth	DuraSpace (Andrew)	2nd quarter	for authentication/authorization
	Build interface	DuraSpace (Danny)	2nd quarter	integrated with Fedora and DuraCloud APIs
DEMO model			12/31 /2012	

## 8.4.1 Test Content

Test	Completed	Details
Set up UVA APTTrust staging fedora repository	✓ July	Set up a fedora repository for AP Trust content and opened access to cloudsyntax.

Test	Completed	Details
Prepare fedora-based UVA content	✓ July	Andrew ingested a book and an image into a local staging fedora repository, including the master images. Mike added AP Trust submission metadata objects linking to that content.
Cloudsync the content to cloud storage (Fedora --> cloud) from UVA	✓ August	Chris successfully initiated CloudSync to copy the content.
Cloudsync the content to the APTrust aggregation repository (Fedora --> Fedora) from UVA	✓ August	Chris successfully initiated CloudSync to copy the content.
Cloudsync the content to cloud storage (Fedora --> cloud) from UNC	✓ September	Chris successfully initiated CloudSync to copy the content.
Cloudsync the content to the APTrust aggregation repository (Fedora --> Fedora) from UNC	✓ September	Chris successfully initiated CloudSync to copy the content.
Cloudsync the content from the cloud storage to the APTrust aggregation repository (Cloud --> fedora)		
Cloudsync the content from the fedora aggregation repository back into the local repository (Fedora--> Fedora, disaster recovery)		
Cloudsync content from a local disk storage to cloud storage and then into the repository.		
Prepare a very large file for ingest (to test >5GB files)		
Cloudsync the large content to cloud storage (Fedora --> cloud)		
Cloudsync the large content to fedora (Fedora --> Fedora)		
Use Cloudsync to restore the large content into fedora (disaster recovery)		

Test	Completed	Details
Cloudsync the large content from the cloud storage to Fedora, updating the datastreams to serve as redirect datastreams to the content in the cloud.		

## 8.5 APT 1.0 Communication Plan

### 8.5.1 Goals

As a part of establishing a trusted infrastructure and workflow, open and regular communication is vital to build understanding amongst partners of current state of APTTrust. We seek to create an open dialog and clearly articulate progress of APTTrust as well as the implications of our architecture on partnering institutions. We also seek to engage our partners in the identification and prioritization of new features and facilitate a more organic and community driven approach toward growth.

### 8.5.2 Audience Definition

The primary audience for APTTrust communications are its member and partner organizations. This community should receive regular updates on the progress of the project, have clear methods to ask questions and gather more information and participate with other members in a discussion of project goals.

The secondary audience for APTTrust communications is the broader academic archiving and preservation community. A vital component of the APTTrust mission is to positively shape best practices and help supply resources to the greater community. To that end we should participate in expanding awareness and dialog about significant developments or important events to the community by using it's communication tools appropriately.

### 8.5.3 Communication Tools

**Primary APTTrust Website** (<http://www.aptrust.org>)

The main APTTrust website should serve as the central resource for external news and non-technical information vital to APTTrust partners. All external communications should be posted here first with links and sharing done as appropriate in other social media tools.

**APTrust on Twitter** (<https://twitter.com/>)

**APTrust Facebook Page** (<http://www.facebook.com/APTrust>)

**APTrust Wiki Site** (<https://wiki.duraspace.org/display/aptrust>)

### Groups & Forums

Discussion groups or forums may be created on topic specific basis as needed and should be listed here along with their purpose.

- **APTrust-tech Google Group** (<https://groups.google.com/forum/?fromgroups#!forum/aptrust-tech>) For discussions by the technical group, closed membership.

**APTrust YouTube Channel** (<http://www.youtube.com/user/APTrust>)

This is for posting additional videos of updates or tutorials. Items here should be presented in the primary website as well. Ownership of this channel is under the [academicpreservationtrust@gmail.com](mailto:academicpreservationtrust@gmail.com) user and not the Google App account since this is not part of the free Google Apps package.

**APTrust Google Apps organization** (<https://www.google.com/a/aptrust.org/>)

### APTrust Email Servers

Email services for the aptrust.org domain are maintained under the Google Apps.

## 8.5.4 Message, Priority of Communication

Under construction

## 8.6 Dec, 2012

---

### 8.6.1 Basic Flow

#### Content

1. Fedora
  - a. UNC
  - b. UVa
2. DSpace
  - a. NCState
3. Filesystem? - **Apr**
  - a. Who will provide content?



## Local to Cloud

1. APT metadata - All content
  - a. What form does this take for transfer to cloud from client?
  - b. What client-side "validation" is required?
    - i. for the AP Trust submission metadata
      1. If we can provide a simple XML schema, this would be straightforward. -ajs6f
    - ii. for policy violations (like invalid replacements of DPN content, versioning issues, etc.)
    - iii. for efficiency (to avoid resynching content that hasn't changed)
  - c. How do we expect the APT metadata to be supplied?
    - i. See 1.b.i. -ajs6f
2. Cloudsync - Fedora content
  - chunking
  - md5
3. DSpace Curation Tool - DSpace content
  - a. Push to DuraCloud
  - b. Invoking this will likely not be available via web console
4. SyncTool and/or UpSync - Filesystem content - **Apr**

## IPS

1. Check if Fedora object already exists
2. Create APT metadata objects
3. Copies content from staging to processed duracloud space
4. DSpace transform
  - a. Unzip
  - b. Wrap with Fedora objects

## Cloud to Aggr-repo

1. Cloudsync, invoked by IPS

## Aggr-repo

1. How to validate that content arrived?
2. Solr?
  - a. Perhaps GSearch or Beverly -ajs6f

## Logging

1. Keep in mind throughout flow

## DPN hooks

1. Ensure metadata/flag is captured
2. IPS? triggers on flag
  - a. recognize a piece of content is flagged for DPN
  - b. minimally write a log message indicating flag was noticed
3. Is APTrust ingest metadata part of what gets submitted to DPN? -ajs6f

## 8.6.2 Admin Web Console

### Requirements

1. Weekly meetings to flush out initial set

### Wire frame

### Data sources

1. Applications
  - a. Fedora/Solr
  - b. Cloudsync
  - c. DuraCloud
2. Define details data required
3. Define details of RESTful interactions

## 9 Technical milestones

### APTrust Partner Demonstration Plan

For a more updated project plan see the [attached PDF](#) with the proposed timeline created using ProjectLibre. These are proposed timelines and some are expected to grow. They serve as an initial set of dates to facilitate a discussion about the work schedule leading up to the partner demo in April.

#### Phase I Milestones set in September 2012:

Task	Lead	Target	Done
Modify Cloudsync to process very large files <ul style="list-style-type: none"> <li>✓ <a href="#">Chunking</a></li> <li>✓ <a href="#">Checksumming</a></li> <li>✓ Test ingest from DuraCloud to APTrust Fedora</li> </ul>	Andrew W  Mike	12/30/12	
Develop Ingest Processing Service <ul style="list-style-type: none"> <li>Code to copy file from staging to new cloud space</li> <li>Log ingest workflow steps</li> <li>Preliminary test of DSpace content through IPS</li> </ul>	Mike	12/30/12	
Develop Admin Interface <ul style="list-style-type: none"> <li>Web application</li> <li>✓ Wireframes</li> <li>Solr</li> <li><i>Export functionality</i></li> </ul>	Andrew W Danny Jeremy Mike Mike	3/1/13	
Client-side tools <ul style="list-style-type: none"> <li>Command line tool for non-Fedora ingest/validation</li> <li><i>Export functionality</i></li> </ul>	Mike	2/1/13	

Test all basic preservation functions of agg repo (submit, verify, audit, <i>retrieve</i> ) for:	Team	4/1/13	
<ul style="list-style-type: none"> <li>Fedora content</li> <li>DSpace content</li> </ul>			
Demo APTrust Phase I functionality at partner meeting		4/29/13	

Note: tasks in italics are planned but may not be finished by target date

# 10 Workflow Description Summary

---

*This page is under heavy construction and working toward Draft Status.*

- [Preservation Package Workflow](#)
  - [Packaging](#)
    - [Fedora Content](#)
    - [DSpace Content](#)
    - [File-Space Content](#)
  - [Staging](#)
  - [Ingestion](#)
  - [Export](#)

## 10.1 Preservation Package Workflow

---

This describes the overall workflow involved the APTrust Preservation Package life-cycle. All creation, updating and deletion of Digital Objects, Metadata and Files for preservation in APTrust is carried out at the originating organization to avoid multiple points of confusion in item curation.

### 10.1.1 Packaging

Partnering institutions must identify content they wish to preserve in APTrust, provide [Package Submission Metadata](#) for those items and prepare a package that can be transferred to the staging server for preservation. Specifics for package preparation may vary depending on the originating system of the preservation item and summarized below. For more information see the [APTrust Package Object](#) documentation.

Adding or updating Items to a package i

Deletions of specific items from a package will be carried out by an explicit Delete option in package metadata but the package itself will always be maintained to allow auditing needed for proper preservation.

### Fedora Content

**Note:** *Packaging of Fedora Content is current undergoing proof of concept testing.*

Institutions should create a Package Object in their local Fedora repository with RDF based references to Fedora Objects to be contained in that package in the RELS-EXT. Packages are staged using the Fedora to Disk feature in Duraspace [Fedora CloudSync](#). Further notes can be found in the [APTrust Package Object](#) documentation.

## DSpace Content

**Note:** *Packaging of DSpace Content is being discussed in preparation for proof of concept testing.*

For content that originates in a DSpace repository we ask that partners prepare items preservation packages by using the [AIP Backup and Restore](#) feature introduced in version 1.7. Additionally a package manifest file will be created with the same name as the related AIP zip file to contain APTrust Package metadata, pointers to the related AIP file and other important information. The current working idea is that APTrust Packages will be transferred to staging using the [DuraCloud Sync Tool](#).

## File-Space Content

**Note:** *Packaging of file-space content will be addressed in more detail in the near future.*

The current working suggestion is to handle File-Space content with a [BagIt](#) Strategy and moved to staging with the DuraCloud Sync Tool as with DSpace Content above.

### 10.1.2 Staging

### 10.1.3 Ingestion

### 10.1.4 Export