APTrust Fall 2016 Meeting Backgrounder Report

Since our spring 2016 working-group meeting at Johns Hopkins University, the Academic Preservation Trust has joined, formally or informally, collaborative projects on relevant topics across the nation and beyond, and our technical staff has nearly completed a full revision of the underlying software that manages the digital deposits of our members.

Technical Progress

The work led by Andrew Diamond with Christian Dahlhausen and Kelly Cobb is a direct result of last fall's experience of our first deposits at scale. Those deposits revealed bottlenecks in high-volume ingest and impediments to accurate diagnosis of where problems existed, none of which had emerged in earlier testing. The good news was that all the safeguards to ensure that no content was actually lost worked properly, but ingest took too long and was too cumbersome and mysterious.

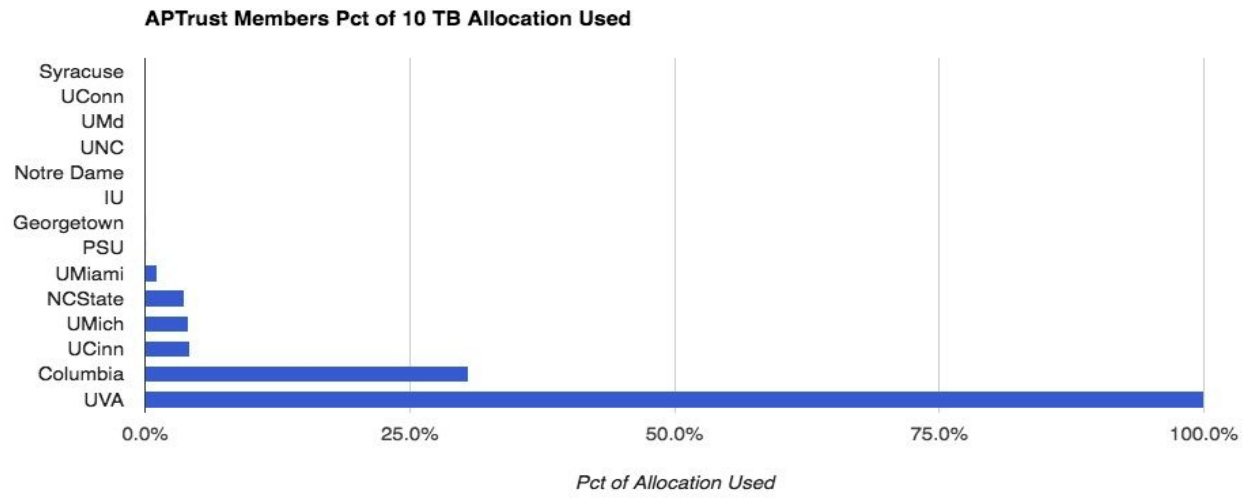
Andrew has reconceived the core software using programming languages and tools that speed the work while ensuring more transparency for diagnosis of issues as well as easing revision and problem resolution. These changes also bring the collateral benefit that they make sophisticated reporting (slated for focused development this fall) much easier and quick to build. Christian has analyzed our compute and storage systems and has employed automation tools to ensure that they can be scaled dynamically to meet fluctuating demands in an efficient, cost-effective manner, with active monitoring so that we can address issues rapidly and with improvements in security. Kelly has concentrated on eliminating components that are no longer needed in Andrew's new design of the applications (and preparing for our migration to Fedora 4) and on the depositor-interface components that will serve as the foundation of major improvements in both general administrative reporting and user reporting on all aspects of the environment. If all that sounds like concentrated, heads-down work, that's just what it was. The work would have been completed before the fall meeting, except that Andrew's efforts had to be diverted frequently to development work for DPN.

For the first half of the 2016 calendar year, our obligations as a DPN ingest and replication node meant that Andrew had to devote much of his time to working with (and often leading) the DPN node-based developers in refining all of the early DPN development work when content started to enter the DPN system, partly through APTrust with the University of Michigan as the our first member-depositor. As with APTrust, deposits at scale exposed issues that no amount of smaller-scale testing would have uncovered. More recently, DPN's leaders, especially Dave Pcolar, were able to negotiate additional technical help from several institutions, most notably Stanford, which had lost its original, full-time DPN developer about a year and a half ago. The new full-time developers will gradually free more of Andrew's time from that project, allowing him (once DPN's version 2.0 of its core software is in production) to turn his attention more fully to the APTrust needs.

As deposit plans for APTrust from our members are becoming clear (and they predict larger deposits later this fall), the timing of our technical development efforts seems to align with those plans well. We anticipate that Andrew's revision of our underlying software will be complete soon after the fall meeting, and focus will shift to expected jumps in deposit volume while Kelly develops and puts into production a suite of depositor-facing reports. With that sequence in mind, a portion of the second day of our fall meeting will be devoted to defining and prioritizing the reports our members need. Through all of this, Christian will help us ensure that we're building our technology and our processes conscious of our goal of achieving certification as a trusted digital repository at the earliest opportunity.

Current Contents

As of September 15, our production environment contained 17.3 TB of ingested content. Here is the distribution of deposits as a proportion of the 10 TB allocation available to each institution included in their membership fee:



Although not visible as bars in this chart, Georgetown and Penn State have smaller volume deposits in the production environment.

Engagement of and with our members

As with many collaborative initiatives across higher education, people from our member universities mostly engage explicitly with APTrust in the form of short-duration, high-intensity interactions, in between longer periods during which they have to pay closer attention to their primary institutional assignments. What makes APTrust different from, for example, the Hydra community or other technical development projects is that our mission is not strictly technical development:

The Academic Preservation Trust (APTrust) is an innovative consortium committed to the creation and management of a sustainable environment for the digital stewardship of academic and research content.

The keywords for exploring this point are “sustainable environment for digital stewardship.” Clearly that environment has major technical aspects, and those have been the primary focus of our work in the early days. Our progress has been consistent and beneficial, and promises much more value in coming years as use increases. But what makes what we do truly sustainable is the our collaborative culture and ongoing collective consideration of the challenges that digital stewardship presents. To say this another way, if we limit our focus to our technical implementations, we will let slip from our grasp the chance to use our collective voice to influence how the world thinks of safeguarding human cultural and intellectual heritage when its form is digital.

I’ve learned in my time with APTrust that we have incredibly insightful people in all aspects of digital preservation, from creative technical developers to masters of content subject-areas and description to legal experts to deans with the wisdom that comes only from deep understanding of the stewardship responsibilities of research libraries, combined with the practical requirements of finding sustainable ways to pay for meeting those responsibilities. At the same time, I observe us mostly talking to each other about the work, even on our own campuses. Pending a positive sense-of-the-Board at our Fall meeting, I’ll be suggesting that we make this a significant area for strategic development for us, taking various forms, each of which involves collaborative development by our members:

- a constantly evolving set of speaker tools that help all members of APTrust, including our deans, inform broader communities about the importance and effective practice of digital preservation for scholarly products/research and for cultural heritage. This would include an accumulating set of case-studies/examples that illustrate what we’re doing and why in more relatable terms than our generic counts of terabytes of content do.
- a most-valuable resources list with numerous topic categories to help newcomers to digital preservation gain knowledge and to help practitioners keep their knowledge fresh, on both practical and theoretical levels, as contributed by our APTrust members..
- a digital platform for individuals from APTrust members (and perhaps guests) to publish their views and to stimulate further debate and, ultimately, action.
- a limited, strategically selected program of affiliations/memberships through which APTrust gains a seat at the table for important conversations. The first of those has been the National Digital Stewardship Alliance, with Bradley and Christian as our representatives. At the upcoming meeting, Bradley and I will propose that APTrust pay for a membership in the UK/Ireland-based Digital Preservation Coalition that recently opened its membership beyond its previous geographic limits. As noted below, we’ll also discuss other affiliations at the Fall meeting.
- and/or other ideas that emerge from our discussions at the Fall meeting.

I’m hoping that members will find this aspect of our mission an interesting one to pursue further. Any work we do on our original tasks and any new ones will benefit greatly from major improvements to our APTrust communications platforms, from our public-facing website to our more operationally focused wiki. That is a subject that has received close attention from our

Communication and Documentation work group. We'll hear a report from that group at the Fall meeting, as we will from two other member-groups, the Trusted Digital Repository work group and the Bagging interest group.

Staff representatives attended the following conferences since the APTrust Spring Meeting last April (APTrust presentations noted with *):

- Open Repositories - June 2016 (Andrew)
- Software Preservation Network - August 2016 (Chip)
- ArchiveIt - August 2016 (Chip)
- SAA - August 2016 (Chip)*
- iPRES - October 2016 (Bradley)
- Internet Archive Library Leaders - October 2016 (Bradley)
- PASIG - October 2016 (Chip and Andrew)
- DLF/NDSA - November 2016 (Chip and Bradley)*
- Proposal in process with DPN nodes: CNI - December 2016 (Chip)*
- Notified of invitation to come: ASERL - December 2016 (Chip)*

Our current list of active partnerships and strategic opportunities include:

- Internet Archive (active)
- SHARE/ Center for Open Science (in development for APTrust, although several of our members are quite active)
- Ontario Coalition of University Libraries (OCUL) (we'll discuss at Fall meeting with representatives)
- Digital Preservation Coalition (DPC-UK/Ireland +) (we'll discuss at Fall meeting)
- Oracle and other potential backend service providers (initial conversations with Oracle continue)
- Software Preservation Network (we'll discuss at Fall meeting with Jessica Meyerson)
- Educopia (we'll discuss at Fall meeting with Katherine Skinner)

Persistent challenges that deserve discussion at the Fall meeting include, but clearly aren't limited to, these:

- Defining digital preservation in practical, relevant terms
    - Unclear distinctions between digital preservation and storage/backup
- Confusion institutions face about choices between services
- How much processing is enough to be ready to digitally preserve?
- Funding preservation activities as core services and not by carving from already stretched collection or technical-infrastructure budgets

Thanks for your attention, and we're looking forward to seeing you at Georgetown!

Chip