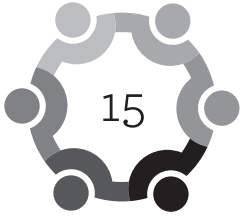

In Medias Res

An Examination of Work in Progress at the Academic Preservation Trust (APTrust) Consortium

R. F. (Chip) German and Kara M. McClurken

It was the summer of 2011. The Academic Preservation Trust,[1] a spark of an idea to create "a robust and sustainable solution for digital preservation,"[2] began as a conversation among research library deans and directors in a meeting convened by James Hilton and Karin Wittenborg of the University of Virginia.

From its beginning, the Academic Preservation Trust (APTrust) was designed to be a collaborative, adaptive, partner-driven model for sustainable, long-term digital preservation solutions. Its creators always intended it to have multiple roles: (a) a digital preservation repository with its own national/international identity and membership; (b) an ingest and replicating node in the concurrently developing Digital Preservation Network, which is a distributed network of national-scale, higher education-run preservation repositories; and (c) a collegial, welcoming forum for surfacing and developing strategies to address the challenges of digital preservation of the scholarly record, research data, and cultural heritage on a massive scale. APTrust's founders believed that individual, smaller-scale solutions to the problem of preserving important

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

digital scholarly and cultural materials could never meet the demand generated by those materials' explosive growth.

As is common with grand ideas, the current state of APTrust reflects the evolution of thinking over the intervening years. Perhaps the most important change has been an expansion of APTrust's initial identity, from a primarily technical project—a collaboratively operated technical environment for digital preservation services—to something more valuable—a consortium of manageable size and scale that connects library leaders and their specialists in working together on the full range of digital preservation issues.

This chapter, written as a firsthand report by two participants in APTrust's development, will explore the consortium's particular organizational history and approach in ways that we hope can inform strategic decision-making about digital preservation. We describe many unresolved issues (and we have few clear answers), but APTrust is providing practice-based experience that may help us all see possibilities to pursue.

Resilience

A glance at APTrust's history[3] reveals the most important observation of all about it: resilience. After the first five years of the consortium's existence, the primary people who conceived the idea and led its early development (we will refer to them as "Gen1") were no longer active in the project, because of retirements[4] and moves to different jobs.[5] Those who succeeded them ("Gen2") inherited many fundamental design concepts about APTrust and have added some of their own, but APTrust has continued both as a technical environment and as a conversation-convening consortium at a steady pace, learning from its accumulated experience and adjusting its course accordingly.[6]

APTrust's Gen1 leadership recruited the initial membership, refined the mission, built the financial structure around commitments from participating institutions, and set the technical direction of APTrust. There were changes in membership, with a few institutions leaving (mainly because some, such as Stanford University, had helped to get APTrust started, but did not expect to participate on an ongoing basis). Others joined during the same time, with the total number of institutions remaining relatively stable at sixteen, depending on when the count was taken.

Anyone who has worked on inter-institutional projects knows that they usually rely on a durable commitment of membership and continuity of leadership to maintain their momentum, or even just to survive. The fact that

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

APTrust continued apace through significant events in those areas suggests that its core concept was a strong one, but we conclude that there are two essential components to APTrust's success thus far: (a) stable, ongoing funding, most importantly and obviously from the University of Virginia (UVA); and equally importantly, (b) a core group of committed individuals/institutions that are willing to take a leadership role in order to ensure that this endeavor does important things.

The significant financial and infrastructural investment by UVA was key to getting the consortium started and maintaining it during some significant transitional periods. In 2014, then-dean Karin Wittenborg and her deputy university librarian, Martha Sites, made the case to then-UVA provost John Simon that the university needed to rise to the challenges of digital preservation, not only for stewarding its own burgeoning digital holdings, but to help find solutions to worldwide needs driven by the huge growth in digital scholarly content and cultural heritage materials everywhere. They argued that an ongoing commitment of funding to APTrust would represent just such a step into a regional, national, and international leadership role. Simon supported the proposal, and in spite of the fact that all three of these key players are no longer at the university, UVA continues to contribute 54 percent of APTrust's operating revenues, with the rest coming from consortial dues paid by other member institutions and APTrust's fees for specific services.

This strong financial base gave members confidence that APTrust would weather the cycles of personnel change that occurred, natural as they were, in the early years of the consortium's development.[7] By providing investment not just in UVA's own digital infrastructure, but also by committing to support the infrastructure of a network larger than itself, UVA has provided the assurance needed to sustain the consortium in times when others might have faltered.

But we would only be telling part of the story if we attributed the resilience simply to relative confidence related to finances. The other key factor is the commitment on the part of UVA and a core group of institutional members to provide human resources as well as financial ones. Even as personnel have changed at many of the member institutions, there continues to be a consistently sized group of dedicated individuals (just about sixteen, interestingly the number of member institutions) who are invested in making the consortium work, in spite of the many other demands on their time and sometimes in spite of their own changing roles and responsibilities.

The membership of this group has changed marginally over time, but it is largely populated with specialists who have accumulated significant experience in the APTrust context. They test new releases, collaborate with APTrust

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

staff to troubleshoot problems and to refine workflows, examine policies, and work towards the creation of a trusted repository. A common characteristic they share is a sense of consortial mission, which helps them move beyond their individual institutional interests to work together and achieve far more widespread benefits.

The Evolution of Technical Design

In the later stages of the first generation of APTrust's leadership, technical development dominated the consortium's activity, and some original plans for the shape of the technical environment changed, as shown in the time line of figure 15.1.[8] Perhaps the most important characteristic of APTrust's preservation repository environment to emerge from the technical planning phase was the decision to use cloud-based Amazon Web Services (AWS) for its digital storage and computing infrastructure. Then-technical lead Scott Turnbull and contributing technical specialists from member libraries linked this decision to another critical one that also remains at the core of APTrust's technical environment today. Whenever possible, APTrust develops its own management software, rather than relying on tools provided by AWS, so that the consortium can move its content to another service provider and adapt the management software to that new environment more simply and rapidly. In addition, APTrust conducts its own fixity checks every ninety days, independent of services provided by AWS, in order to ensure data integrity. These mitigations of dependence on a vendor's tools are defining concepts of APTrust.

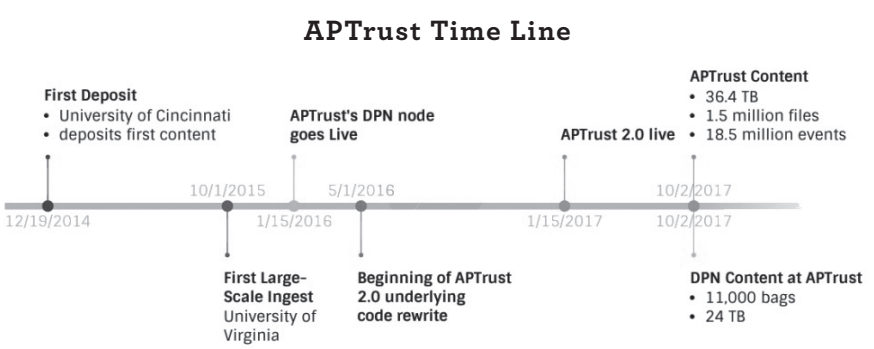

Figure 15.1 • APTrust time line

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

The first generation of APTrust's leadership also resolved to limit the scope of its initial technical environment to bit-level preservation in a "dark" (meaning not designed for easy or frequent access) repository environment. This close focus enabled the technical staff to concentrate exclusively on fixing the ingest-management code they had discovered to be slowing ingest at the high volume that APTrust was expected to handle. Even with this concentration, the cure took time. To fully fix the problem, current technical lead Andrew Diamond began a complete rewrite of APTrust's original management software in 2016 (completed in early 2017), based on lessons learned from early production experience. Portions of APTrust's ingest process now operate 100 times faster than they did with the old code.

From the earliest days, APTrust's Gen1 leaders saw it not only as a consortial entity in its own right, but also as a node of the Digital Preservation Network (DPN). DPN's history deserves its own chapter, and we will not try to relate it here, except to note that much of APTrust's technical development effort was tightly linked to DPN's own effort. DPN's evolution as an organization, its own refinements of its character, and its technical complexities required consistent time and effort from the APTrust staff in ways that were important and valuable to the entire digital preservation ecosystem, but it also had the collateral effect of slowing down APTrust's development by stretching the same staff resources over two major projects. The effects on APTrust's schedule were not mirrored, however, in financial costs. The DPN project fully reimburses its nodes for any services they provide. Our lesson from this experience is that, should the need for DPN development require resources from APTrust in the future, we will supplement our staffing to cover it with temporary or part-time employees in order to avoid adverse effects on APTrust's own momentum.

APTrust had to ask and answer a long sequence of questions about exactly how it would function as a DPN node. Among them, APTrust originally considered offering deposit services for non-APTrust members to DPN, but it subsequently chose instead to refer nonmembers to deposit through the DuraCloud Vault DPN node (a partnership between the University of California at San Diego and DuraSpace). Although APTrust remains a primary DPN replicating node, and is available to house copies of content from all DPN members, it only serves as a DPN deposit node for APTrust members. In doing so, APTrust can serve as an important service partner in the DPN network, but it still can focus its work on the APTrust core mission.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

The Art of the Conversation—A Consortium at Its Best

Although everyone associated with APTrust has experienced regular instances of high-quality, digital preservation-themed conversations at many levels throughout the consortium's history, both member representatives and staff did not originally explicitly articulate this benefit as a primary one of APTrust participation. We believe that the value of these conversations comes from the unusual combination of scale and organizational size that APTrust embodies. It is working on the global challenge of scale: the gap between the explosion of digital scholarly and cultural heritage content and the adoption of effective preservation practices for that content. The consortium has simultaneously limited its membership; this is a conscious choice that promotes the development of close relationships within a manageably sized group of geographically dispersed professionals (unlike professional associations, whose sizes can make personal connections uneven).

The focus on a complex global challenge by a congenial community of moderate size enables the rapid alignment of ideas about how to work meaningfully on pieces of that challenge. The work of APTrust is connected to the participants' daily work, but it places that work in a much larger frame of reference. The consortium meets face-to-face periodically (currently twice a year) in pleasant venues at relatively low cost and in a meeting group whose size (unlike national conferences) is small enough to promote good personal interactions, while at the same time it is large enough to contain a wide diversity of roles, experiences, and institutional contexts.[9]

In the last several years, the APTrust meetings have included visitors whose presence has multiplied the value that comes from the APTrust consortium's scale and scope. They have included key representatives from the Internet Archive, the Ontario Council of University Libraries, the Center for Open Science, the Software Preservation Network, the MetaArchive Cooperative/Educopia, the Boston Public Library, and others. Because of APTrust's modest size, the range of roles represented by participants, and their breadth of interests, the meetings have allowed for deep conversations and exchanges of ideas that do not commonly happen in other meeting circumstances.

Financial Foundation and Staffing

The move of APTrust's preservation repository to "production" status in late 2014 required an accompanying "production-ready" budgeting plan (see

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

table 15.1). APTrust staff had by then developed a template for budgeting and financial reporting that included tracking a reserve fund. The approach that APTrust takes to the disclosure of its finances is another defining aspect of its character, with two essential elements: precise cost-allocation and transparency.

One of the advantages of Amazon Web Services, compared as it often is to institutional data centers, is that anyone can see exactly how bills are calculated (including rate structures for the varying components that comprise computing and storage services).[10] When combined with the ability to activate and shut down such components as needed, this detailed-billing practice puts the power of fine-grained control in the hands of people who use AWS. Amazon also provides similarly fine-grained cost-planning tools that allow its users to make close predictions of the costs of different service configurations. Although APTrust reports both cloud-service expenditures and predictions about new service costs at summary levels, staff members are prepared to lead curious members through the details of those costs at any time.

Table 15.1 • APTrust FY16–17 budget report

APTrust Budget Report August 30, 2017		2016-17 ORIGINAL BUDGET	2016-17 REVISED BUDGET	2016-17 ACTUALS	2016-17 VARIANCE
Revenues	REVENUES				
	Ongoing UVA funding	$380,000	$380,000	$380,000	$0
	Sustaining Member dues	$320,000	$320,000	$320,000	$0
	Grants				$0
	Subscription Fees		$8,250		-$8,250
	REVENUE TOTALS	$700,000	$708,250	$700,000	-$8,250
Expenditures	EXPENDITURES				
	Staff costs (includes FB)	-$541,197	-$541,197	-$541,197	$0
	Equipment costs	-$6,000	-$6,000	$0	$6,000
	Travel costs-APTrust	-$30,000	-$30,000	-$31,940	-$1,940
	Event costs	-$25,000	-$25,000	-$432	$24,568
	Consultants/Temps	-$48,000	-$48,000	-$20,475	$27,525
	Cloud technology costs	-$35,000	-$35,000	-$59,837	-$24,837
	Other technology costs				$0
	Misc service costs				
	Memberships	$0	-$3,500	-$3,154	$346
	DPN reimbursements	$60,000	$60,000	$37,215	-$22,785
Net Expenditures	EXPENDITURE TOTAL	-$625,197	-$628,697	-$619,821	$8,876
Operating Positive/Negative Balance	OPERATING BALANCE	$74,803	$79,553	$80,179	$626
	TRANSFER TO/FROM RESERVE			$80,179	
	APTrust Reserve Balance	$785,199	$785,199	$865,378	
NOTES:		Date of this report: August 30, 2017 FY runs July 1 to June 30 Cloud technology costs include short-term extra expenses for compute during code rewrite			

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

APTrust operates "lean." Its director and associate director work only half-time on APTrust. They split their work between the consortium and other duties in the UVA Library. Three full-time specialists comprise the technical staff, who perform or coordinate with member specialists the development of APTrust's repository environment. They are responsible for its ongoing operation, and they explore, recommend, and—when recommendations are accepted—develop new services and capabilities.[11]

At professional conferences, the staff and representatives from member libraries present information about how APTrust works and what the consortium has learned from its growing experience. Staff travel to member institutions to better understand their needs, and they visit organizations doing similar work in order to share information. Consequently, conference registrations and travel by staff comprise a significant budget category. Member institutions pay separately for their representatives' travel. What is not a significant budget category is advertising. APTrust does not pay to sponsor anything but its own activities, which include its face-to-face meetings. Those meetings are usually held at member institutions at much lower cost than other meeting-location options. Meeting at the institutions also allows more host institution-based staff to engage with APTrust, an important collateral benefit.

Challenges to (Consortial) Digital Preservation Efforts

APTrust production experience is pointing the way to the next big questions to explore, and sometimes to new services that the consortium may want either to provide directly or to otherwise make easier for those practicing digital preservation to obtain. Experience is also highlighting problems that APTrust and the digital preservation community in general need to address.

As we will discuss later in this chapter, memory institutions regularly cite financial challenges as major barriers to digital preservation efforts. From APTrust's experience to date, we can say that cost inevitably becomes a significant complication that hinders long-term planning for digital preservation at our member institutions. That noted, we must add a caveat: no one has paid "extra" for high volumes of content deposited at APTrust. Our 16 members each get 10 terabytes of deposited digital content included in their member dues. To date (2017), two have reached that quota, with one more member institution close behind. Most others have much to deposit before they will reach the quota. So costs alone cannot explain why there is not more content in

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

APTrust. From our informal conversations with other higher education-based digital preservation projects, we believe that APTrust's growth in deposits over time, shown in figure 15.2, are substantially above the average for such projects, but when compared to the national worldwide scale of *need* for digital preservation, that growth is still far short of sufficient.

What is impeding the growth of preservation deposits? Although we know that worry about cost is obstructing long-range planning for some of our institutions, other factors are stalling deposits from institutions that have enough quota "headroom" to put significant volumes of content into the repository before incurring additional costs at APTrust. Those impediments, based on anecdotal conversations with our members and confirmed in part by recent surveys,[12] include

- Uncertainty about what content is important enough to receive the "special" preservation action that APTrust represents
- Institutional backlogs of material that have not been processed to the point that digital preservation staff are prepared to deposit it (including factors such as incomplete metadata)
- APTrust-specific ingest requirements that are sufficiently cumbersome that institutions are finding the work difficult to prioritize against other uses of staff time

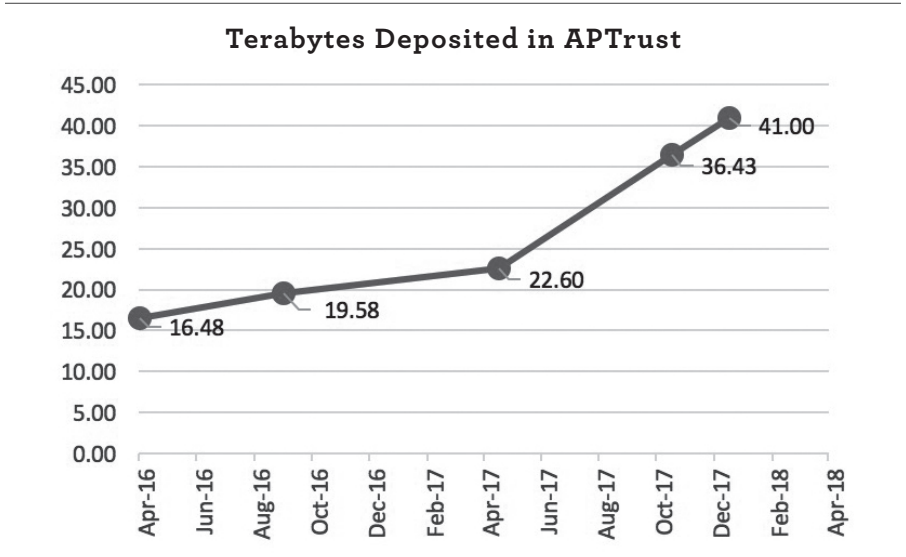

Figure 15.2 • APTrust deposit volume through November 2017

The fact that APTrust represents an outside, unintegrated element in an institution's current digital preservation workflow and requires extra work that feels disproportionate

Too many preservation strategy choices (even for APTrust members) without deeply compelling reasons to invest effort in one over another

When a challenge is huge, and when no single approach appears to be the answer, relative paralysis should not be a surprise. For most of the APTrust institutions that are depositing significant volumes of materials, digital preservation staff are doing what they can in the direction of what they think may be right, and APTrust is planning to make its services easier to use and more flexible. Plans for the near future include

Expanded service offerings for different levels of assurance (including a significantly lower-cost, off-line-only option without independent fixity checking, but with a choice of AWS data centers and perhaps additional platforms, beyond AWS)

A specific service for published digital materials (such as e-books for which a library has subscribed to perpetual access, but wants to ensure this) when alternatives for such preservation are not present

A menu of choices that consolidates additional deposit processes, where members can check boxes for additional deposit beyond APTrust to such locations as DPN and the Digital Public Library of America

Opportunities for the aggregation of content across institutions. The HathiTrust Research Center provides an outstanding model for other entities to copy in building a large corpus of similar materials that can become the subject of comprehensive research. APTrust is exploring how to assist in the development of such bodies of content and how to make available to researchers the powerful computing resources with which to study them cost-effectively.

Easier deposit mechanisms, including drag-and-drop deposit that does not require specialized "packaging" such as the creation of content packages in the BagIt format (APTrust's current standard). Such mechanisms will also include API-integration with commonly used, open-source content preservation tools.

Access for members to workspaces where they can apply cloud-based format migration tools to their content

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

Trusted Digital Repository (TDR) compliance: APTrust has always intended to operate a Trusted Digital Repository, and work should be completed soon to document the repository's compliance with relevant standards. A question under active debate as we write this chapter is whether or not the added value of going through formal certification as a TDR is worth the cost (a cost that adds to the cost of preservation).

To Consortium or Not to Consortium

A common belief of APTrust participants is that a consortial approach to the challenges and opportunities that digital preservation presents to research libraries should work, can work, and is working—mostly—with APTrust. But might other solutions provide a better approach? The alternatives to banding together to work collaboratively on such a challenge are (a) to solve the problem within each institution, (b) to buy a solution from a vendor, or (c) to pretend that the problem does not exist until someone else solves it for you. The fact is that one solution will not fit all institutions, all are in use, and we hope that choice (c) disappears completely as a strategy.

The Shrinking University Data Center

One part of APTrust's core design creates an alternative to a dependence on university data centers and is present in some other digital-preservation consortium designs. To those of us who have been responsible for running university data centers, it has been common practice in the past to build services on the so-called marginal capacity of our centers. This theory assumes that institutions are building storage and computing capacities into their data centers for many reasons, and that adding some of both for a use such as digital preservation only requires "marginal" investments, rather than creating new, dedicated infrastructure for such uses. Two major, related trends call this traditional practice into question. Under the financial pressures facing higher education in general, many institutions are seeking ways to hold down the costs of their data centers. One strategy that is growing (rapidly, we argue) is to move commodity services away from increasingly specialized university data centers and toward massive cloud-based suppliers, which are seen by some analysts as providing them more reliably and cheaply.[13] Often, data-at-rest and

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

the computing power needed to ingest and manage it (the usual components of digital preservation) are seen as commodity services.

Even when university data centers continue to provide such services, the more frequent use of cost-allocation models to divide costs according to use means that those centers will likely bill their users at their full proportional costs for both technology and staffing. For some university data centers operating under such cost-allocation models, users discover that the "fully loaded" cost is more than that of cloud-based providers when the services are directly comparable.[14] APTrust was designed to provide a transparent way for member institutions to prove this point right or wrong, a role it is performing well for its members.

Academic Consortium, Not Vendor

APTrust actively works to avoid a danger that is present for other academic consortia which develop production services: the risk of being seen as a vendor, rather than its true nature, which is an investment in collaborative problem-solving that operates directly within the missions of institutional members. These are fundamentally different approaches. A higher education consortium that is designed to use the affordances of collaboration to solve complex problems ultimately aims to work itself out of existence. It does so by solving the complex problem it was invented to address, by doing its best until a next wave of approaches replaces it, or by living just long enough to prove its irrelevance to the ultimate solution.

Vendors are not invented to go out of business. They are invented to generate dependence on the products they create and sell, and to continue that dependence by adapting their product lines to maintain relevance and value to their customers, whose needs change over time.

Both vendors and academic consortia have important roles to play in digital preservation, precisely because they do different things for different reasons. For a consortium such as APTrust, it is never advantageous for its identity to be confused with that of a vendor. Three elements are major contributors to such a misperception. The first is the legal framework. The more extensive the legal language is that describes the relationships between parties in the consortium, the more it feels like a vendor relationship to institutional members. This is a major reason why APTrust's leadership has not pursued an independent legal identity for the consortium. Instead, it operates as an activity of the University of Virginia Library, its host location. APTrust specifically uses the language of a consortial alliance to document its relationship to its members.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

For the same reason, the APTrust consortium assumes no risk by accepting digital content deposited by its members. Instead, it champions a collaborative approach that reduces each institution's risk when that institution deposits content in the shared preservation repository. The consortium further reduces institutional risk through succession planning (i.e., what happens to the deposited content if the members decide to terminate APTrust) and by the firm establishment of a reserve fund to guarantee operations for a transition period, if ever needed. At a simpler and more immediately practical level, the consortium reduces risk by implicitly inculcating in its members, through relationship-building among the consortium's individual institutional representatives, a sense of common commitment to the protection of each other's content.

A second element contributing to the danger of being perceived as a vendor is the requirement for billing. In order to ensure that each institution covers its fair share of the costs of operating a preservation repository, bills are generated based on usage. APTrust's approach to this element is extreme transparency, as with all of its finances. Members can and must be able to dissect bills completely down to their cost basis. Such an approach is rare, if not entirely absent, with vendors.

The third major element in the consortium's danger of being perceived as a vendor is that it has employees, a fact that can sometimes inject a border of "otherness" between institutional representatives and a consortium's staff. APTrust self-consciously acts to eliminate this border both structurally and behaviorally. At the structural level, UVA's role as host and as the main funder of APTrust intentionally reduces the financial burden borne by other institutional members for supporting staff, while ensuring that such staff are seen for what they are: UVA Library employees.

At the behavioral level, APTrust specifically creates a staff culture that promotes a sense of deep-rooted collegiality between its staff and member-colleagues. The success of such an approach will occur when member institutions regard the APTrust staff as extended members of their own staffs, and when they rely on the consortium staff to do things they choose not to pay their local staff to do. By this measure, APTrust is only partially successful, due in large part to the intensive technical-development phase of work (including for DPN) that has occupied most of the APTrust staff to date. In the future, the APTrust staff will be able to contribute more visible and wider-ranging expertise that can help institutions better develop and implement strategies for digital preservation on a large scale.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

Collaboration of Collaborations

Although work on DPN may have slowed APTrust's momentum somewhat—unintentionally on everyone's part and due largely to how APTrust chose to staff for it—this work on DPN has provided lasting benefits as well. Most important among them is the establishment of durable, productive relationships between the staffs of a wide range of higher education projects working in digital preservation. The projects are diverse, but they are united by a common belief that the digital preservation of scholarly and cultural materials is already an essential resource for the future intellectual development of human society. Even when combining our efforts, we represent only a small fraction of the capacity needed to meet the needs and challenges of this work. There is room for all of what we are doing and so much more that competition between us is an irrelevant concept.

As a group that now includes projects and services beyond the DPN network, we have drafted a statement of shared values that articulates the things uniting us and that documents our shared commitment to the body of digital content that we collectively steward. The statement includes an agreement in principle, when one of our repositories is at risk, to play helpful roles in ensuring that the content involved is not lost. To many of us, this notion of diverse, distributed preservation linked by goodwill and a common dedication to preserving human knowledge is the key to the kind of preservation model that can be effective in the future.

Clearly such a concept is not contained within national borders, and APTrust has sought to establish digital preservation alliances beyond the United States, such as applying for (and being accepted for) membership in the U.K.-based Digital Preservation Coalition. We live in a time when daily examples remind us that digital scholarly knowledge, research data, and cultural heritage are susceptible to politically motivated eradication, as much as they are to technical failures and human error. International capabilities are key elements of mitigation for that risk.

On another axis related to issues that cross national borders, who will ensure that what we save today will be meaningfully accessible in the future? Both file formats and the software that renders data usable are at high risk for obsolescence and disappearance. When we at APTrust do the mental calculations of how much effort it will take to continually convert old materials into something we can use at any one time, we think that effort will drastically outpace any resources available to support it (people and money). We are intrigued by the idea that instead we need to refine our capabilities to preserve the software

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

that can work with the old data, and by doing so sustain our ability to derive from it usable versions as needed over time.

To that end, APTrust, on behalf of its members, has been an active participant in the development of the Software Preservation Network (SPN)[15] and expects to play a helpful role in software preservation as new opportunities emerge from SPN, which is another example of a collaboration of collaborations.

More Dollars and Sense

As we noted earlier, anyone working on digital preservation has heard the most often-cited impediment to doing it routinely: we cannot afford it. This can be literally true, as our experiences with APTrust have demonstrated. If you have large digital objects (such as video) and you have petabytes of them, as does one of APTrust's members, Indiana University at Bloomington, the costs of digital preservation at any reasonably rigorous level of assurance are staggering.

But in what context are they staggering? As we write, APTrust's core service (assuming that its funding base of member dues and UVA support remains consistent) is billed to members who wish to deposit more than the amount included in their base membership fee at a cost of $420 per terabyte per year. Given that the annual member dues and UVA support go nearly completely to personnel expenditures, the $420/TB/year is a good representation, when excluding personnel,[16] of the cost of APTrust's core, high-assurance preservation service.

For that cost, APTrust provides preservation storage copies in three separate availability zones in Amazon's near-line storage service on the East Coast and copies in three separate availability zones in Amazon's off-line storage service on the West Coast. APTrust conducts independent fixity checks on the content every ninety days (a cost-intensive step that is under significant debate in the Digital Preservation Coalition's online community as we write).

There are many points worthy of further discussion, including what kinds of individual digital objects (huge or not) need that many copies and that much geographic and technical diversity, and, if not, what would be the cost of what is truly needed. To shorten the point, is less preservation assurance for a lower cost sufficient, especially when we are considering quantities of material at this volume? To begin to answer this question requires thinking about comparable efforts between digital and physical objects. In order to focus the question a bit, we will assume that the comparable objects are both rare and important to our institutions. What do we spend to protect a comparable measure of physical objects?

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

We do not intend either to use complicated formulas previously published or to invent a new one in our attempt to approximate an answer.[17] Instead, we will use a case example from the University of Virginia. Anticipating the complete renovation of the university's primary library building, Alderman Library, the university in 2017 began the process of building a second library storage facility on the edge of its grounds, which is intended in part to house some rare materials for the long term. The construction of Ivy 2, as it is known in this planning phase, includes a dedicated capital budget as well as some related funding from other university sources for preparatory activity, and it will require ongoing operational funding. The capital outlay is $7.9 million for 20,000 gross square feet (and ceilings more than thirty feet tall). A rough estimate of the building's capacity is three million volumes.

Simple math reveals that the capital expenditure amount is $2.63 per volume. If we assume that the life span of the building (without major capital infusions) is fifty years, that translates to a capital cost of $0.053 per year per volume. Operating costs need to be added to that number. Utilities costs for Ivy 2's older sibling (the original Ivy Stacks, smaller and with a less precisely managed climate-control system) were $24,345 for fiscal year 2017, or $1.92 per square foot per year. That would be $0.013 per volume, at Ivy 2's density. Other facilities costs are allocated at UVA on a square-foot basis, most recently at $8.75 per square foot per year. That would be roughly $175,000 per year for Ivy 2, or $0.058 per year per physical volume.

Adding these cost components together, in this model (which, as in the APTrust example, excludes operating-personnel costs) the total cost for physical volumes is $0.071 per year per volume for one copy stored in one location. In these terms, how might we compare a physical volume to a digital one? The most obvious way is by using an average size for the digitization of a volume. For HathiTrust when it does future budget and storage-space planning for high-quality digitized materials, the organization assumes approximately seventy megabytes per volume (it is lower for older volumes, but for the purpose of this discussion, we will assume the number for newer volumes is the most helpful one).[18]

Using APTrust's fee for high-assurance preservation storage and HathiTrust's average size of a volume, a digitized volume could be expected to cost $0.028 per year[19] to preserve, with three copies each in two distant areas of the country and 90-day fixity checks, excluding personnel. If we regard one data center as roughly equivalent to one physical repository location (such as UVA's Ivy 2), then the comparable digital cost would actually be half of that—$0.014 per year per data center. The comparison is then $0.071 per year

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

per physical volume against $0.014 per digital volume, by this rudimentary method, excluding personnel.[20] In short, it is reasonable to conclude that it is much less expensive to preserve a single digital object than it is to preserve the physical form of a similar object.

Given this conclusion, why are costs such an impediment? First and most obviously, they largely represent additions, not substitutions. We will discuss this further later in the chapter.

More practically and immediately of interest may be this observation: part of the problem is how money is allocated and by whom. A new building is generally allocated from a different source and on a different time line than annual expenses. For example, the money for our off-site storage building was paid for by capital funds and approved as a line item in the university's overall multiyear capital budget for submission to the Commonwealth of Virginia by the university's governing body, the Rector and Visitors.

The greatest costs of physical storage come from long-term capital investments; the greatest costs for digital storage, especially for those dealing with petabytes of data, are incurred and budgeted annually. Their funding comes from the UVA Library's operating budget and competes with other library services to be included in that budget. More importantly, the costs for the other library services, for the most part, are not going away even as digital preservation needs and associated costs grow.[21]

How do we reconcile the differences in the ways that physical and digital preservation's primary costs are framed in concept and funded in process? Should we seek some form of capital investment or periodic large lump-sum funding for digital preservation, and if so, how do we frame that concept as an alternative (or supplement) to annual operating budgets? If we had a means to gain access to an amount similar to the $7.9 million set in the UVA case for the preservation of the physical collections, we could create an endowment that would generate income that would substantially reduce the cost-impediments to digital preservation at scale. When donors give us their digital and physical collections, should we ask them to contribute to this kind of endowment, providing ongoing support for digital preservation storage as well as physical storage? If every donor of relevant materials who gave money for processing/supplies also gave money to a digital-preservation storage endowment, we would move closer to funding those needs.[22]

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

The Long Tail Challenge

At the Coalition for Networked Information's membership meeting in December 2016, Elliott Shore, the executive director of the Association of Research Libraries, challenged a panel of representatives of digital preservation projects (including APTrust)[23] about organizational schemes that required members to pay significant amounts of dues to participate. His implication was that we were limiting digital preservation to the elite research institutions; smaller libraries and other memory organizations were held back from participating by high member dues ($20,000 per year per institution in APTrust's case).

Shore's point was a valid one, and not new to anyone on the panel. Despite the justification we gave at the time of the importance of leading the way in developing strategies for digital preservation at scale, the "long tail" issue comes up regularly in a wide array of forms. For APTrust, this emphasized a dilemma that drove the consortium to a creative notion, consistent with its character: what about offering each member the ability to serve as a content hub in order to facilitate deposits from other entities?

Each member is able to decide for itself if and how it wants to play that role. In some instances, a library may choose to simply categorize as "public service" any costs it incurs in serving as the member-depositor for content supplied by an entity such as an area historical society. Alternatively, it could charge some amount for serving in that role to the entity—the amount could be simply the incremental cost of depositing that entity's content. At UVA, this idea matched easily with the university library's evolving interest in helping area entities with digital content and little wherewithal to preserve it.

The Really Big Picture

Many points raised in this sketch make our conceptual frameworks about preserving future digital scholarly objects and cultural heritage materials bafflingly complex. It is clear that there are sufficient differences between most of what we have preserved in the past and most of what we will want to preserve in the future to limit some of what we can learn from past experience. If we simply faced the prospect of preserving digitized versions of printed volumes (in fact, we do face that prospect or something similar over the short term), we would be able to say some simple things: digital preservation is comparable, and may be ultimately cheaper, per volume than physical preservation, and yet we are still not prepared for, or planning adequately for, the fact that those costs are

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

additions to our current preservation expenditures. They do not substitute for current expenditures, unless we decide to abandon existing efforts to preserve physical objects, selling the buildings that house them to generate more cash for the cause of digital preservation (an idea we are not recommending). This is a major issue today that is distracting us from deeper, longer-term thinking. Our institutions are not grappling directly or effectively with even today's issue (yet).

Perhaps an even more significant issue facing us is the growth in size of a common digital object, which is one unit of measure for digital preservation. If we assume that digitized books, as noted above, represented a common digital object for which we were planning capacity, that object might be expected to measure 70 megabytes. If we project that in a few years the common size of a digital object which is a candidate for digital preservation may be that of today's 4K feature movie, it could measure 15 terabytes (up 375 percent from the last 2K digital generation of this kind of content).[24] If the average digital object to be preserved evolved over time from the size of a digitized book to the size of a 4K feature movie, that would involve a more than a 215,000-fold increase in the storage required per object (this is not a prediction, but an interesting mental exercise to contemplate).

For APTrust even today, some member institutions are reporting that they have already entered this future. Similarly daunting is the trend to preserving full research data sets to support the reproducibility of original results and reuse to do new analysis. So another complicating factor to comprehensive solutions is the astounding growth of the average size of digital objects and bodies of data, a multiplier that has to be applied to the rapidly growing number of objects that are candidates for preservation.

APTrust's experience is starkly demonstrating the central and (so far) unresolvable conflict of digital preservation for the future: the explosion in the volume of digital materials reflecting human intellectual and cultural development versus the lack of money to pay for their preservation. William Kilbride of the Digital Preservation Coalition noted current growth estimates that succinctly highlight the problem in a presentation at the 2017 meeting of the Preservation and Archiving Special Interest Group at the University of Oxford:

Data Creation:	Approximately 60 percent per annum
Storage Capacities:	Approximately 40 percent per annum
IT Budgets:	Approximately 2 percent per annum[25]

Many others have described this issue over time. If we cannot identify funding to pay for meaningful preservation at a scale proportional to today's production

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

volume of digital scholarly objects and cultural heritage materials, systematic digital preservation will not happen.

This has happened before in the cultural community. The size of paper-based collections grew exponentially in the 1950s. And we are still dealing with the challenges of housing all of those paper-based collections today.[26] But there are a few differences here. For one thing, paper is not as ephemeral as digital forms. The stewards of paper objects often are willing to store their paper until it reaches the end of its useful life. We cannot wait that long to take preservation action with digital objects: technological obsolescence, the fragility of the physical media on which they are stored, the challenges of accessing password-protected information after its creator's death, and the ease with which one can inadvertently delete such objects makes addressing the digital explosion more urgent than the last explosion of (paper) document creation.

From our perspective, the biggest risk is not that we will choose unwisely when winnowing down candidates for preservation. It is that the human effort needed to select what is likely to be most important in the future will be overwhelmed by accumulating masses of unassessed digital materials. Without sufficient funding to preserve more than we can actively select, we are likely to lose significant volumes of material that are important to future intellectual progress, creating what some observers describe as memory holes. And epistemological Alzheimer's at the scale of human civilization is a chilling prospect. What exactly is the definition of acceptable loss in this context?

As we write this in 2017, APTrust's community is in the midst of a full review of its accomplishments to date and directions for the future. While admitting that APTrust has contributed usefully to the evolving ecosystem of digital preservation by answering some questions and framing new ones from its practical experience, our perspective is that there is clearly more, higher-order work to be done. The consortium must be deeply engaged with colleagues worldwide on the most massive issues (such as the explosion in data creation as well as the corresponding lack of human, monetary, and infrastructure resources to deal with that exponential growth) that stand in the way of comprehensive digital preservation.[27] Those issues exist at levels that are much more fundamental and enduring than a short-term quest for specific technical solutions that reduce the rapid growth of digital preservation costs.

Of course, we cannot be paralyzed by the monumental proportions of the challenges that face digital preservation. We must, like our most productive members operating within their institutional contexts, continue to do what we can in the direction of what we think may be right. That said, self-congratulations need to wait until the biggest challenges have become solvable.

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

The mission of the APTrust consortium, and that for all of our colleagues working on similar efforts, must be grounded in the global context, and our work has just begun.

NOTES

1. Academic Preservation Trust, http://aptrust.org.
2. Martha Sites, "The APTrust Story," conference proceeding, Association of Research Libraries, Chapel Hill, NC, 2013, p. 1, http://tinyurl.com/y9ps75hg (https://doi.org/10.18130/V3SF2P).
3. The primary source for much of the early history of APTrust is Sites, "The APTrust Story."
4. UVA Library Dean Karin Wittenborg and UVA Library Interim Dean Martha Sites retired in 2014 and 2016, respectively.
5. UVA Vice President and CIO James Hilton (2014), APTrust Program Director Suzanne Thorin (2014), and APTrust Technical Lead Scott Turnbull (2015) moved to other jobs.
6. APTrust's governing board has experienced less change, with three of the seven voting-member seats having turned over (one has done so twice), all due to retirements.
7. This is a two-edged factor, of course. Dependence on continuing UVA funding could cause hesitation on the part of depositors at other institutions, who are uncertain of the durability of UVA's funding commitment. We have not uncovered any direct evidence that deposits have been constrained by this concern, and the UVA has repeatedly described its funding support as ongoing and permanent.
8. For example, the original technical design anticipated relying on Duraspace's services and the Fedora repository architecture, but subsequent decisions moved APTrust's "back end" to Amazon Web Services, with Fedora managing content metadata. In more recent changes, APTrust is moving from Fedora dependence to a Fedora-friendly environment (and is still defining exactly what this last phrase means).
9. Active participants in the meetings have titles that include (but are not limited to) repository managers, preservation administrators, digital preservation librarians, developers, digital collection librarians and administrators, and library deans and directors of information technology.
10. This is not exclusive to the Amazon Web Service. We are aware that other cloud services offer this level of detail.
11. In these early years, the two largest categories of expenditure in APTrust have been the cloud-service costs and staff, as indicated in regularly published budget reports in which viewers can see the other, smaller cost categories as well.
12. For example, a survey including 170 responses was the subject of a preliminary report at the Preservation and Archiving Special Interest Group (PASIG) conference at Oxford University on September 13, 2017. The presentation used for that report is here: Evviva Weinraub and Laura Alagna, "Beyond the Repository: Integrating Local Preservation Systems with National Distribution Services," figshare, http://tinyurl.com/yc2fnu4l (https://doi.org/10.6084/m9.figshare.5415136.v1).

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

13. There are many views on this question, and comprehensive analysis is complicated. See, for example, ECAR-TCO Working Group, "TCO for Cloud Services: A Framework," *Research Bulletin* (Louisville, CO: ECAR), April 24, 2015, http://tinyurl.com/y9ryxbsv (https://web.archive.org/web/20171127160628/https://library.educause.edu/~/media/files/library/%202015/4/ewg1503-pdf). Yet in the 2016 version of an annual Campus Computing Project survey, director Kenneth C. (Casey) Green reported at that year's EDUCAUSE Annual Conference that higher education IT budgets have not fully recovered from the Great Recession, and at the same time 42 percent of institutions are expecting to spend more money on cloud computing. Casey Green, "Key Campus IT Issues: Personnel, Instruction, Budgets, Security, and Analytics," Campus Computing Project, October 2016, http://tinyurl.com/ycrqgj6y (https://web.archive.org/web/20171127160226/https://static1.squarespace.com/static/5757372f8a65e295305044dc/t/586ff909db29d6ae7c8663b8/1483733267591/CampusComputing2016-EDUCAUSE+Presentation_0.pdf).
14. This has been the observation of the University of Virginia Library in budget planning for the 2017–18 fiscal year.
15. For more information on the Software Preservation Network, see www.softwarepreservationnetwork.org/.
16. We are excluding the personnel costs because they are the hardest to compare, with many staff having mixtures of duties that make them difficult to specifically allocate. Instead, we are focusing on other specific components of cost to do rough comparisons between physical objects and digital objects and to highlight costs related to the evolution of size of some common digital objects.
17. For more on the costs of storing a physical book, see Paul N. Courant and Matthew "Buzzey" Nielsen, "On the Cost of Keeping a Book," in *The Idea of Order: Transforming Research Collections for 21st Century Scholarship* (Washington, DC: CLIR, 2010), 81–105. For more on the costs related to digital storage see Richard Davies, ed., "The LIFE² Final Project Report," http://tinyurl.com/yd8hmp5h (https://web.archive.org/web/20170926185913/http://discovery.ucl.ac.uk/11758/1/11758.pdf).
18. Sandra McIntyre, HathiTrust, "Re: Quick Data Question," e-mail to Chip German, APTrust, August 29, 2017.
19. The calculation is $420/1,048,576 ($0.000400, the per-megabyte per year cost) × 70 MB (size of digitized volume) = $0.0280 per volume per year in two geographically distant data centers.
20. Many caveats apply: the costs for physical preservation are likely to be significant underestimates in this method, and those for the digital include two high-expense factors (at least): many more copies for the price, and active monitoring of the integrity of the digital object that does not often occur with the physical one. The physical-object preservation scenario may actually be more comparable to one of the new services APTrust is developing: a Glacier-only, one-data center service. That would be priced at $60 per terabyte per year, leading to this calculation: $60/1,048,576 ($0.000057, the

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).

per megabyte per year cost) × 70 MB (size of a digitized volume) = $0.0040 (cost per volume per year in a single data center).

21. Even in areas where we may perform services less often (like binding physical journals), the cost per volume has gone up, meaning that we cannot reduce expenditures in such areas in order to fund digital storage.
22. There are many other related notions: if we calculated the costs of the preservation-storage of digital objects in the purchase prices of the digital collections, then some money could come from collections budgets (which admittedly have their own major pressures to try to balance). Should we ask researchers to factor in the costs of long-term (or even short-term) storage in their grant applications? Combined together, over time, we could shift our thinking about funding sources so that annual preservation-storage fees were less of a hurdle.
23. One of the authors of this chapter, Chip German, was a member of the panel. For information about the other panel participants, see "The Digital Preservation Ecosystem: A Community Conversation with Providers of Services," http://tinyurl.com/ycau23lt (https://www.cni.org/events/membership-meetings/past-meetings/fall-2016/schedule-f16).
24. This estimate in 4K movie preservation-object size is drawn from Slide 8 of a presentation at the PASIG 2017 conference: Mathieu Giannecchini, "The Éclair Archive Heritage Cinema Use Case: Rising to the Challenges of Complex Formats at Large Scale," figshare, http://tinyurl.com/ydh2axbm (https://doi.org/10.6084/m9.figshare.5415025.v1).
25. William Kilbride, "Sustainable Digital Futures," conference presentation, PASIG 2017 Oxford Bootcamp Day, September 11, 2017, figshare, http://tinyurl.com/y7jgkhcg (https://doi.org/10.6084/m9.figshare.5414986.v1).
26. Mark A. Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing, *American Archivist* vol. 68 (fall/winter 2005): 211, http://tinyurl.com/ybcrsxl8 (https://web.archive.org/web/20170926191401/www.archivists.org/prof-education/pre-readings/IMPLP/AA68.2.MeissnerGreene.pdf).
27. Clearly, we are indebted to the many pioneers whose insights are critical to understanding the array of global challenges, notably including for us at APTrust Dr. David S. H. Rosenthal (at this writing, his blog on digital preservation issues and many related subjects continues at http://blog.dshr.org).

From *Digital Preservation in Libraries: Preparing for a Sustainable Future*, edited by Jeremy Myntti and Jessalyn Zoom (Chicago: American Library Association, 2019).