

# Digital Preservation Primer: An Introduction to Digital Preservation at UVa



Lauren Work  
Digital Preservation Librarian



University of Virginia Library  
Preservation Week 2016  
[lw2cd@virginia.edu](mailto:lw2cd@virginia.edu)

Hi everyone - thanks very much for coming today. And happy preservation week!

I'm Lauren Work, and I'm the Digital Preservation Librarian here at UVa. I started in this position in January of this year, so if we have not met yet; hello!

Just as a brief introduction to my place within the library for those who may not be familiar- I'm part of the small but mighty Content Stewardship group of the library, which is responsible for overseeing the physical and digital preservation of collections, with the goal of making them permanently accessible.

Within this group, I'm part of Preservation services, and can be found on the second floor of Alderman Library, in room 202. Please feel free to come say hello.

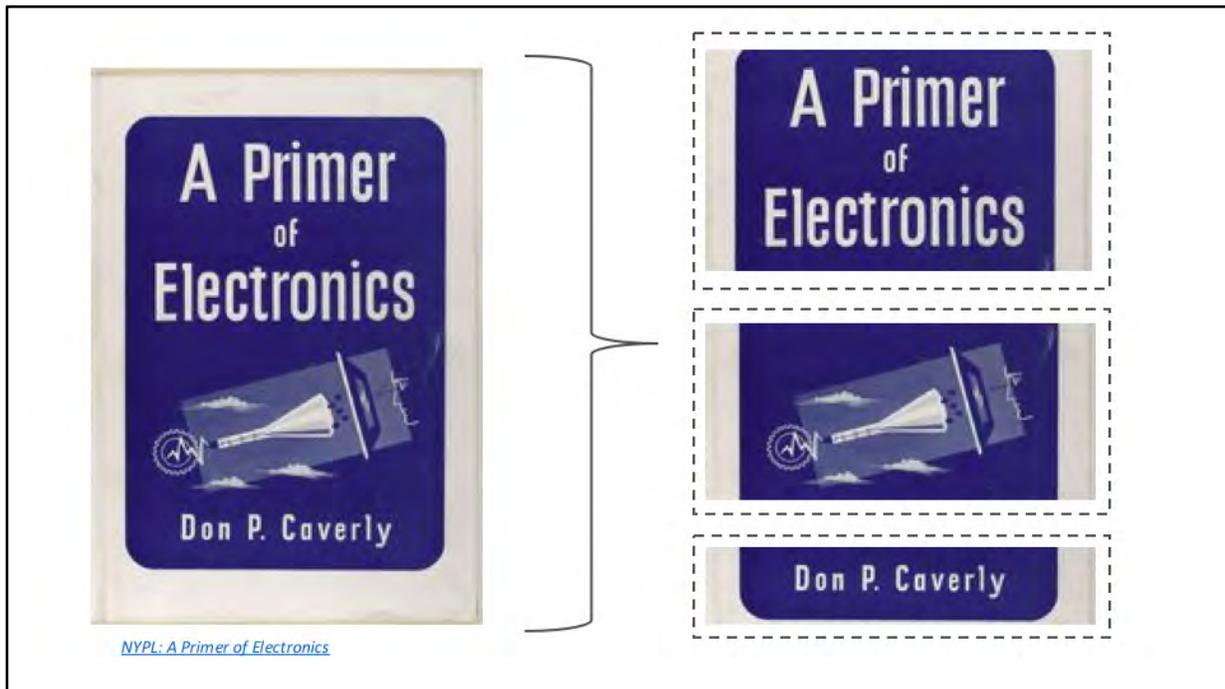


So, what do I do as the digital preservation librarian here at UVa?

I'm responsible for the sustainable preservation of university digital resources, and "digital resources" runs the gamut from websites to archival hard drives and floppy disks, to born digital scholarship.

I personally also see my role as a highly collaborative conduit in the library, connecting and serving the digital preservation needs and workflows that range from Special Collections to software development to Collection Access & Description, to scholarly communication and digital humanities and back again.

So really - much like my coworkers in traditional preservation, I ingest, stabilize, secure and monitor university resources to make them sustainably accessible - I just get to use my hands a bit less, and stare at screens a bit more to do so.



I framed this talk as a digital preservation “primer” for a few reasons:  
First, alliteration is fun!

Second - for those not directly involved in digital preservation work, even within libraries, the subject of digital preservation is kind of a broad, sweeping concept.

And short detail, (such as the detail I just gave about my own position), don’t really offer insight into the *practice* of digital preservation - what does it look like to “do” digital preservation? I aim to break these concepts into component parts and give you some use cases to help increase understanding today.

Lastly, I formed this talk as a “primer” because I think it’s key to understand the concepts that form the basic building blocks of digital preservation, as ultimately it allows for better understanding of why and how we approach important digital preservation projects here at UVa.

My goal is for you to leave here today knowing more about digital preservation as a whole, and with an understanding about how we are undertaking digital preservation work at UVa that ties into the community of concepts and best practices

surrounding digital preservation.

“Digital preservation is the active management of digital content over time to ensure ongoing access”

*Library of Congress*  
<http://www.digitalpreservation.gov/about>

So. Let's start at the beginning.  
What is digital preservation?

The Library of Congress, one of the leaders in the library and preservation field, has this definition:

How many people, based on this definition, would have an idea of what I do all day as your digital preservation librarian?

“Digital preservation combines policies, strategies and actions that ensure access to digital content over time.”

*Association for Library Collections and Technical Services  
Working group  
<http://www.ala.org/alcts/resources/preserv/defdigpres0408>*

Here’s a slightly more detailed definition from the Association for Library Collections and Technical Services.

Ok, now we’re talking about policy and active management as well as access to digital content over time - we’re getting closer to what it means to actively engage in digital preservation.

And I want to pause here for a moment to say - I am running us through these definitions both as an introduction to digital preservation as a concept, but also to provide a quick snapshot of the history of the field.



And I want to provide this history because, in the mid to late 90s, when this version of the Apple computer was the hottest thing going, it was also the time when people in the library and cultural heritage world were really trying to get their heads around what to do with the huge volumes of digital content both being produced and also sitting on obsolete media from scholars and donors on their shelves.

And they were grappling with these definitions and what they meant in terms of policy, standards, stewardship and technological development.

## The Five Organizational Stages

- 1) **Acknowledge:** Understanding that digital preservation is a local concern;
- 2) **Act:** Initiating digital preservation projects;
- 3) **Consolidate:** Seguing from projects to programs;
- 4) **Institutionalize:** Incorporating the larger environment; and
- 5) **Externalize:** Embracing inter-institutional collaboration and dependency

[Kenney and McGovern, 2003](#)

Strides were first made on the administrative & policy end, as some had argued, successfully, that the biggest impediments to digital preservation were organizational, rather than technical.

So approaches from that time in the aughts included these Five Organizational Stages for Digital Preservation from Kenney and McGovern, the Trusted Repository Audit Certification, and other high level, workflow, policy and strategy related work. All very important.



[Leslie Jones "Action."](#) CC BY-NC-ND

However - much of the progress of these policy and strategy-related approaches did not address what practitioners - those actually responsible for stewarding, ingesting, and protecting the digital content - should actually be doing to take **action** on digital preservation.

<p style="text-align: center;"><b>Action!</b></p> <p><b>National Digital Stewardship Alliance Levels of Preservation</b></p>		Level 1 (Protect your data)	Level 2 (Know your data)	Level 3 (Monitor your data)	Level 4 (Repair your data)
	Storage and Geographic Location	<ul style="list-style-type: none"> <li>- Two complete copies that are not collocated</li> <li>- For data on heterogeneous media (optical discs, hard drives, etc.) get the content off the medium and into your storage system</li> </ul>	<ul style="list-style-type: none"> <li>- At least three complete copies</li> <li>- At least one copy in a different geographic location</li> <li>- Document your storage system(s) and storage media and what you need to use them</li> </ul>	<ul style="list-style-type: none"> <li>- At least one copy in a geographic location with a different disaster threat</li> <li>- Obsolescence monitoring process for your storage system(s) and media</li> </ul>	<ul style="list-style-type: none"> <li>- At least three copies in geographic locations with different disaster threats</li> <li>- Have a comprehensive plan in place that will keep files and metadata on currently accessible media or systems</li> </ul>
	File Fixity and Data Integrity	<ul style="list-style-type: none"> <li>- Check file fixity on ingest if it has been provided with the content</li> <li>- Create fixity info if it wasn't provided with the content</li> </ul>	<ul style="list-style-type: none"> <li>- Check fixity on all ingests</li> <li>- Use write-blockers when working with original media</li> <li>- Virus-check high risk content</li> </ul>	<ul style="list-style-type: none"> <li>- Check fixity of content at fixed intervals</li> <li>- Maintain logs of fixity info: supply audit on demand</li> <li>- Ability to detect corrupt data</li> <li>- Virus-check all content</li> </ul>	<ul style="list-style-type: none"> <li>- Check fixity of all content in response to specific events or activities</li> <li>- Ability to replace/repair corrupted data</li> <li>- Ensure no one person has write access to all copies</li> </ul>
	Information Security	<ul style="list-style-type: none"> <li>- Identify who has read, write, move and delete authorization to individual files</li> <li>- Restrict who has those authorizations to individual files</li> </ul>	<ul style="list-style-type: none"> <li>- Document access restrictions for content</li> </ul>	<ul style="list-style-type: none"> <li>- Maintain logs of who performed what actions on files, including deletions and preservation actions</li> </ul>	<ul style="list-style-type: none"> <li>- Perform audit of logs</li> </ul>
	Metadata	<ul style="list-style-type: none"> <li>- Inventory of content and its storage location</li> <li>- Ensure backup and non-collocation of inventory</li> </ul>	<ul style="list-style-type: none"> <li>- Store administrative metadata</li> <li>- Store transformative metadata and log events</li> </ul>	<ul style="list-style-type: none"> <li>- Store standard technical and descriptive metadata</li> </ul>	<ul style="list-style-type: none"> <li>- Store standard preservation metadata</li> </ul>
	File Formats	<ul style="list-style-type: none"> <li>- When you can give input into the creation of digital files encourage use of a limited set of known open</li> </ul>	<ul style="list-style-type: none"> <li>- Inventory of file formats in use</li> </ul>	<ul style="list-style-type: none"> <li>- Monitor file format obsolescence issues</li> </ul>	<ul style="list-style-type: none"> <li>- Perform format migrations, emulation and similar activities as needed</li> </ul>

So, in 2012, the National Digital Stewardship Alliance, an initiative from the Information Infrastructure and Preservation Program at the Library of Congress, set out to change that - with a simple, tech platform-agnostic, modular approach to actually doing the work of digital preservation. It's meant to help guide both those just beginning with digital preservation programs as well as organizations trying to advance their existing work.

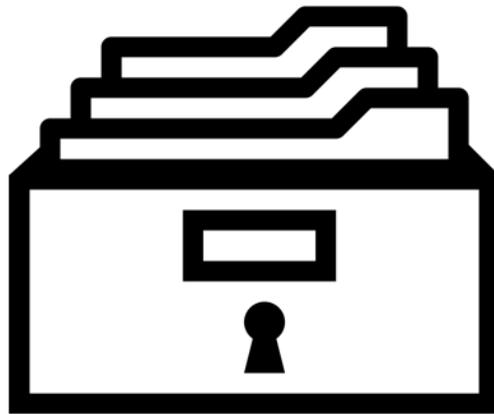
Five major conceptual areas were identified as important areas for action - ways to evaluate what your organization is doing to mitigate risk of loss, and to identify concrete next steps to take for digital preservation.

This is meant to be a living document, with an understand that it is work in progress as the digital preservation field grows and changes, and as technology does the same.

And it helps form the basis of our current approaches to digital preservation, as we'll see in a minute as we go through use cases here at UVA - but I wanted to provide this brief history to show how the work we do comes out of actionable guidance for practitioners and strategists alike. And these approaches are reflected in the tools and policy that have continued to emerge and develop as people do the work of

digital preservation.

## Digital Preservation at UVa



So let's dive into some of the work and planning we're doing here at UVa.

## Born-Digital Scholarship

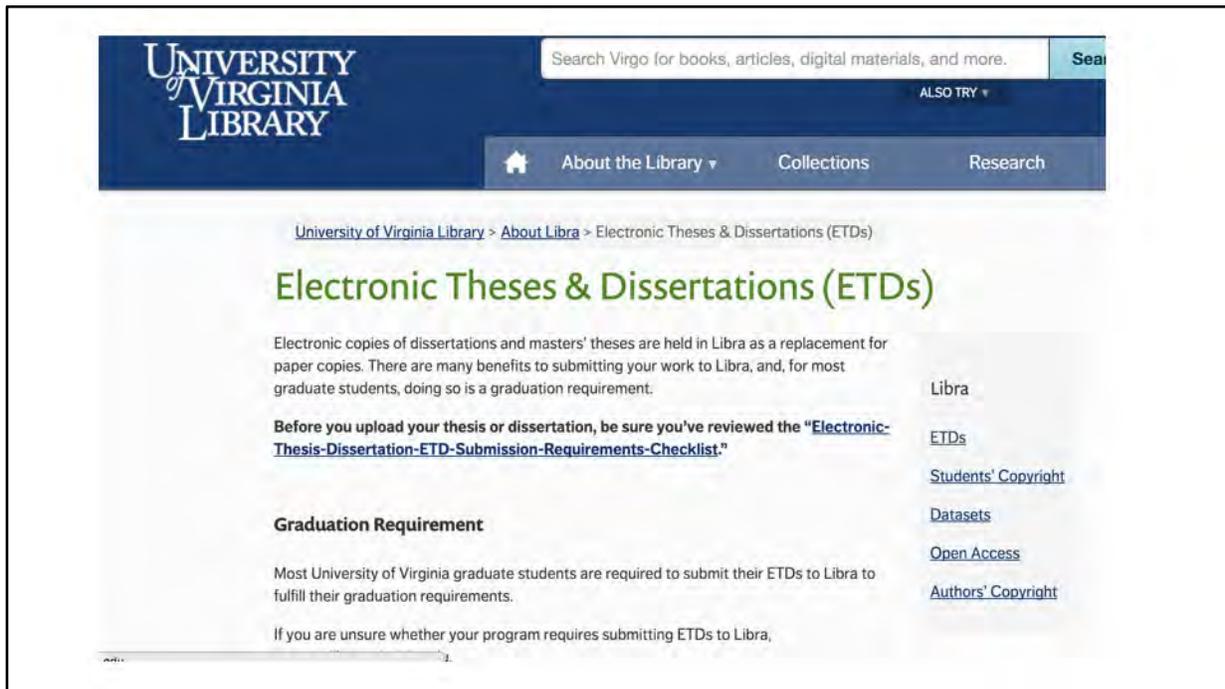


*Graduation by Rediffusion. CC BY 3.0*

Scholarship has continued to take many forms over the years, and I could give an entire talk just on preserving the websites, images, databases, audio, social media, and 3D data that now make up components of modern scholarship. And maybe I will in the future!

But today I wanted to focus on our digital preservation work related to one of the most sweeping changes in academia as a whole - that of electronic theses and dissertations. Many libraries stopped accepting paper copies for deposit and preservation within the last several years, and so preserving digital scholarship for long-term access within an academic repository is of the utmost importance.

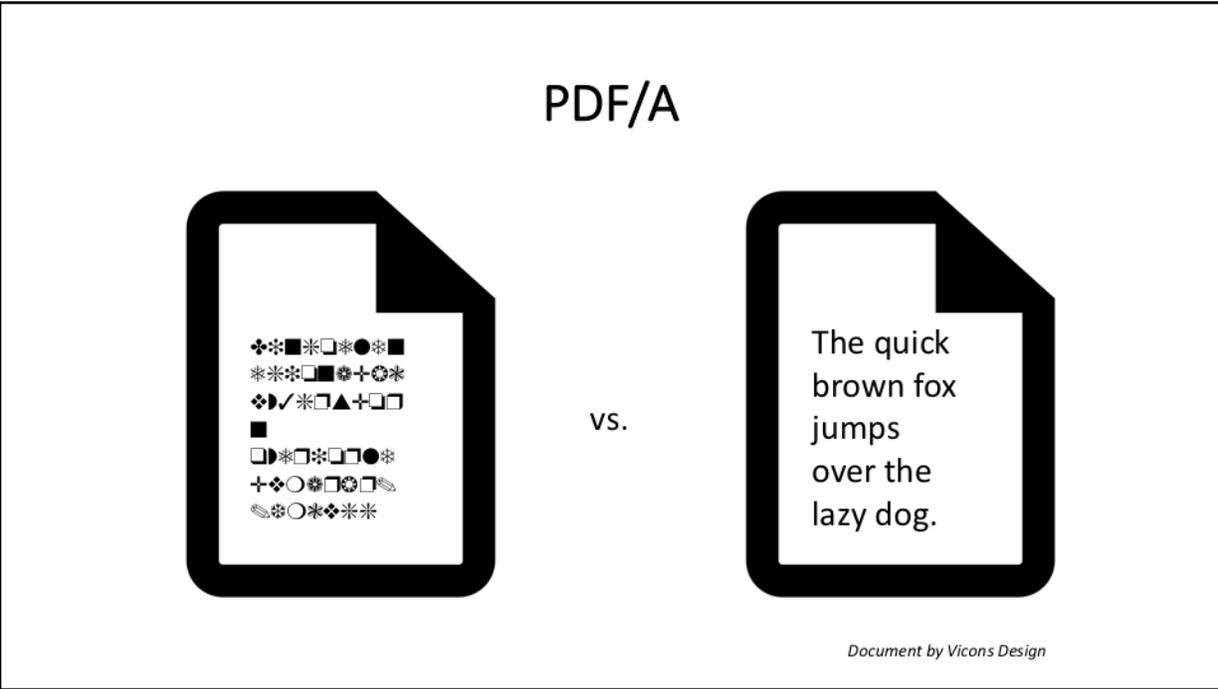
Here at UVa we're currently approaching the preservation of these assets in several ways, but I'll touch on two today.



The first is at a very granular level - that of PDF/A or archival format.

Currently, students submit a plain PDF to Libria, and there are no other format restrictions beyond that. And there are opportunities to provide a more stable, standardized form of PDF that will render more reliably in the future.

We're in the beginning stages, but are working toward establishing PDF/A, as a potential future standard for thesis deposit, which many other academic institutions have done to help better preserve their scholarship.



So what is it? Simply put, PDF/A is an international preservation standard for PDF that is widely adopted as a format. It is device independent, meaning it can be opened on more than 1 type of reader, and is entirely self contained.

Self contained means that the document itself is structured to contain everything needed for an accurate representation of its contents. And it also means no external links to fonts or executable files like Javascript, no video, and no encryption.

The best way to think about this concept is to think about an experience that many of us have had – attempting to open an older document or file that may display broken images, or strange block of text where font should be. PDF/A prevents these things by self-containment, and provides us with better preservation strategies for our scholarly content.

## Academic Preservation Trust



The second approach for digital preservation is at a higher level, with the Academic Preservation Trust.

UvA is the leader of AP Trust, which is a consortial effort of academic institutions to preserve their digital scholarly content. AP Trust pools the resources of its members to build, support, and scale a large, cloud-based repository system. This allows for best practices of digital preservation including geographic distribution, redundancy and recovery of content, fixity checking, storage of metadata, and more.

So these are a few things we're actively working on for the preservation of digital scholarship.

## Digital Preservation of Legacy Media



What about university or donor digital content that only exists currently on unstable media like floppy disks, CDs, or hard drives? What are we doing to preserve this content?

Pictured is an actual example of a floppy disk from Special Collections.

I am sure many of you have something sitting under your desk or in a closet that you are dying to get rid of, or wondering if we even have the hardware anymore to examine these items.

# F.R.E.D.

Forensic Recovery of Evidence Device



Let me introduce to you FRED - our forensic recovery of evidence device, and an important tool for the digital preservation of legacy media from a wide range of sources at the university.

If that sounds like CSI to you - it's meant to. The digital preservation field has found itself dovetailing in some ways with crime scene investigation, in that we aim to securely, and without impact, obtain, examine, and stabilize the contents of media - but for the purposes of long term preservation and provenance, not evidence.

Again, this is similar to my colleagues in traditional preservation - they may receive an item for preservation and notice mold, or an insect. They would identify any issues, and take standardized steps, such as placing the item in a freezer to stabilize & isolate the item. Similarly, I can identify viruses or corrupt files with the FRED, and treat them accordingly to ensure that the content itself is stabilized, and that it cannot harm other works, much like isolating a moldy book.

We've actually just received our new FRED, and I'm very excited to get it up and running, but I thought we could take a quick spin through what this type of digital preservation looks like when you actually get content off of the floppy disk.





Here is our floppy disk again. On the outside label, it claims to be an “archive” from May-Nov of 1990.

To review the actual content on this legacy media, we might go through several preservation steps, including:

An inventory and photograph of the item itself,

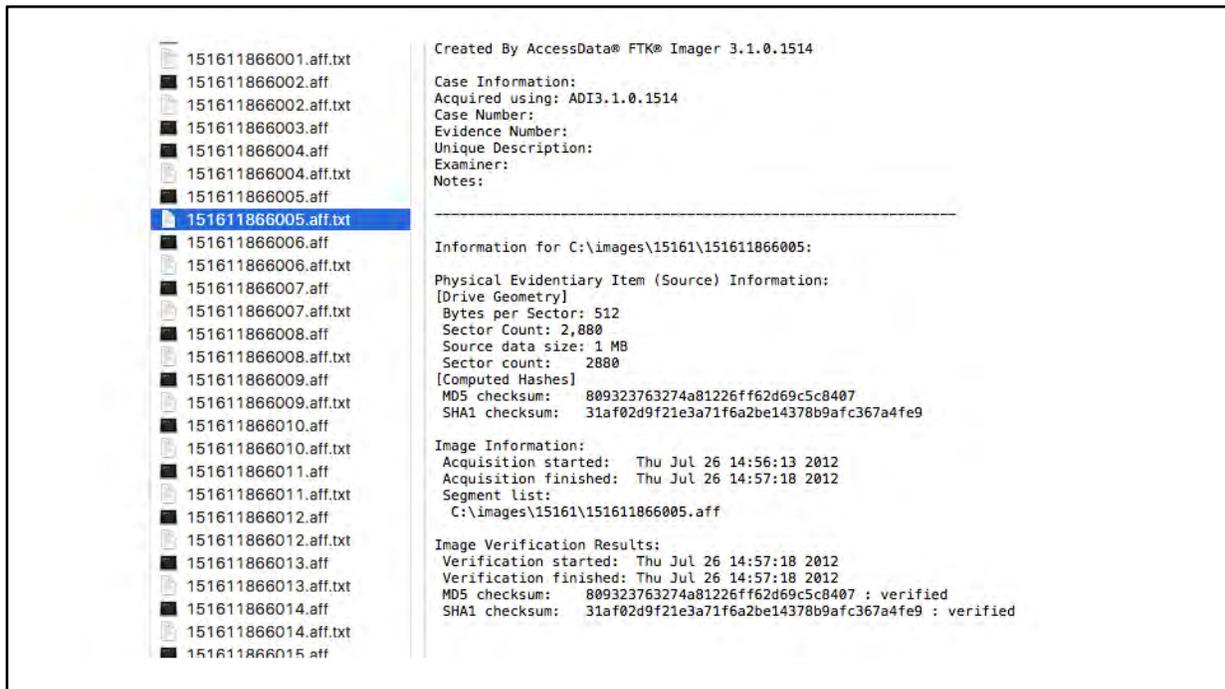
a careful review of the object for damage that could impact both the preservation of the actual object and the ability to read the media on it for digital ingest,

use hardware to insert the physical disk and connect it to a write-blocked port on the FRED, to ensure we don’t accidentally write to or change any files

We would then create the disk image with an open source software called BitCurator,

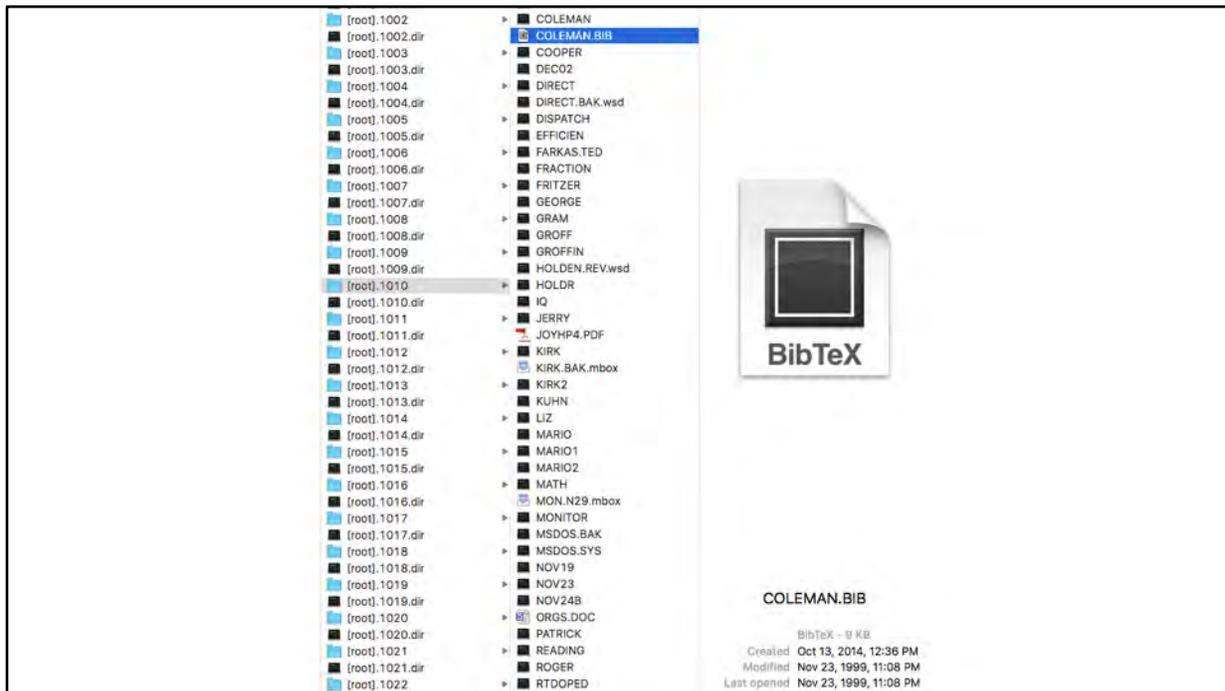
We can also do things like perform a virus scan, run checks for sensitive information such as credit card information or social security numbers, and create checksums to allow us to later validate that the disk image has not changed, and so on.





So you may be wondering - why a disk image? Several reasons. A disk image captures each sector bit for bit on a disk and preserves the entire file system structure. An image also captures all segments of a disk, including deleted files, versions of files, and unallocated space, which can contain other hidden files. So if you simply copied the files over instead of creating a disk image, we might miss other potentially important information.

So we've imaged our disk image - and this is what the result looks like. The highlighted number ending in .aff.txt is the accompanying information file from our disk image, which is the file ending in .aff right about it. The aff extension is an advanced forensic format, which is an open format for the storage of disk images and metadata. What you see in this text file is some of that metadata, including our checksums and some basic technical metadata about the disk.



And here is a closer view of documents that are present on the floppy - you see a PDF, and mbox, which is an email format, some docs, and something called bibtex, which is reference management software.

So this type of digital preservation process allows us to preserve content that by getting it off of legacy media, where it is at risk and inaccessible, allows us to make sure that information is captured, ingested, stabilized and securely packaged in a way that will then allow us to do things like move the digital objects to Special Collections to allow them to apply their expertise for the next steps of the collection, and send the digital objects to our repositories.

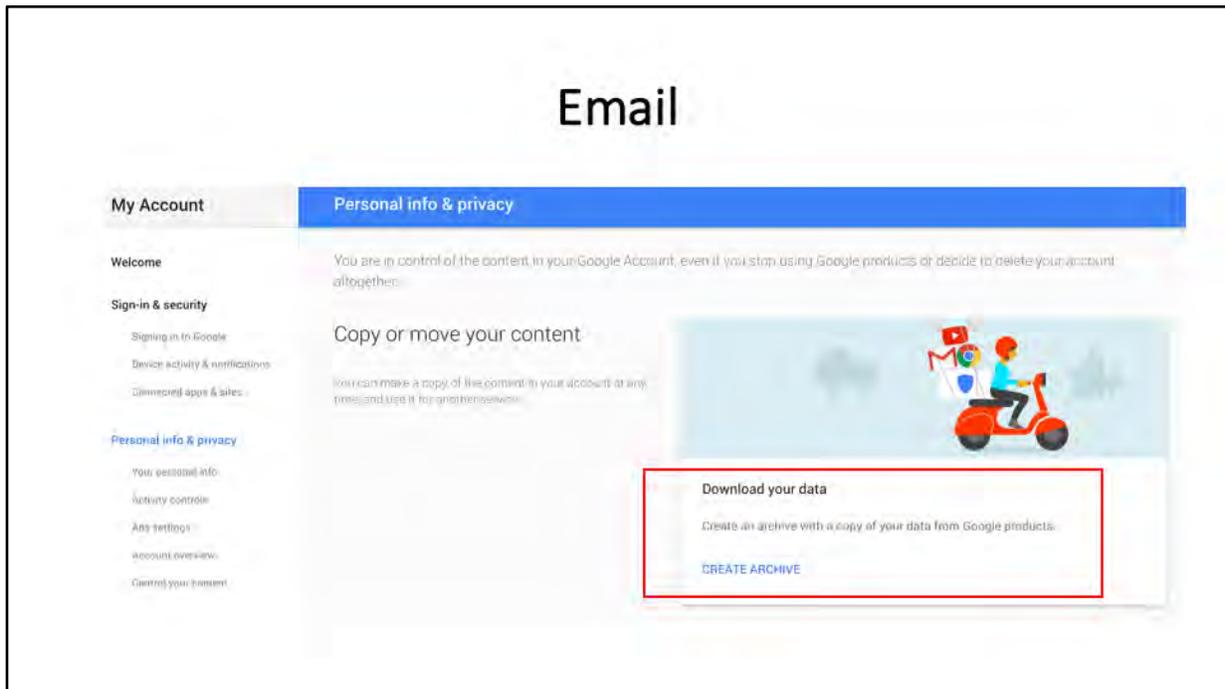
## Digital Preservation of Communication



Created by Aaron K. Kim  
from Noun Project

Changing back to the born-digital part of the preservation puzzle, I want to talk a bit about some of the work we're doing to help ingest, stabilize, and preserve various forms of communication related to the university, as well as how we are thinking about ways to make these records accessible.

As we're all well aware, we now have more ways to communicate than ever before. And the vast majority of this communication is electronic, and some very ephemeral. I'll touch on two types of communication that we've recently been focused on: email and social media - namely Twitter.



We'll start with email.

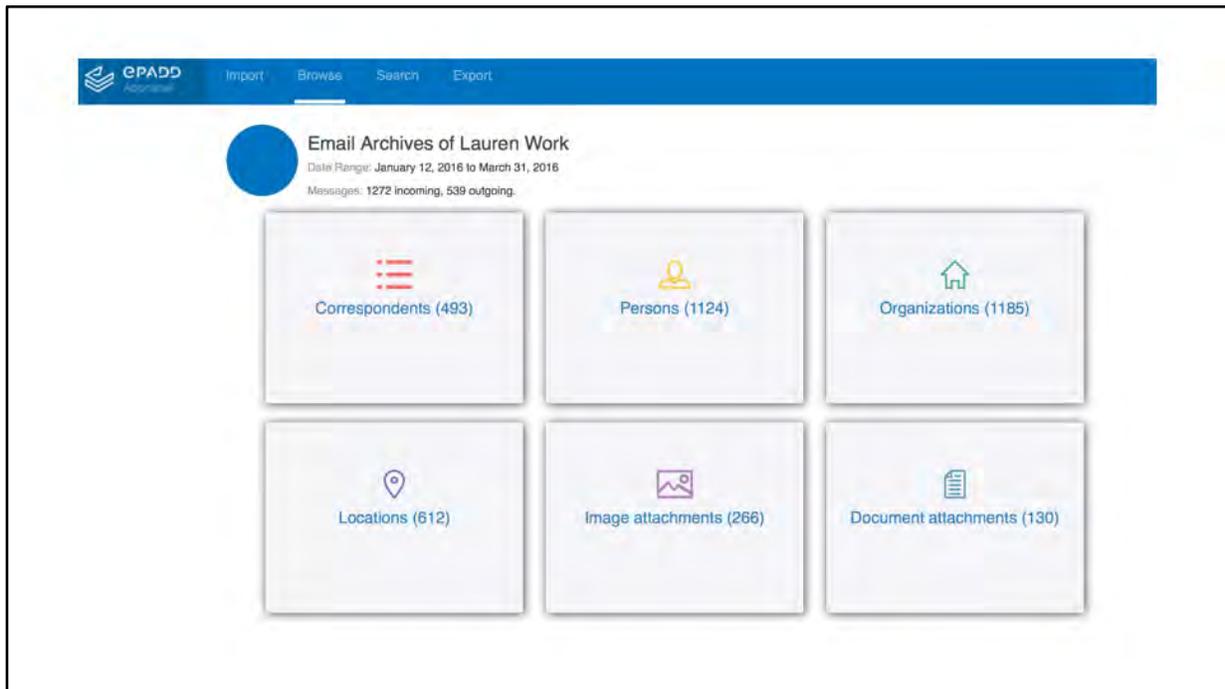
The vast majority of correspondence - be it from writers who plan to gift their papers and correspondence to Special Collections, university presidents fulfilling their duties, or just humble digital preservation librarian emailing with their boss many times a day - all of this happens over email.

And we are responsible for the preservation of this content when it comes time for a donor or president to deposit their correspondence. While printing emails may have been desired or necessary in the past, there are several ways to approach the digital preservation of email messages now.

The first, and one of the most basic, is shown on the screen now. Downloading your own email archive - this is an example from Gmail. Has anyone done this, even for backup? You can get your calendar, contacts, emails, etc. And this is one way to have people transfer their emails for preservation

But there has also been new developments in the tools available that ensure the secure delivery, and preservation of, correspondence





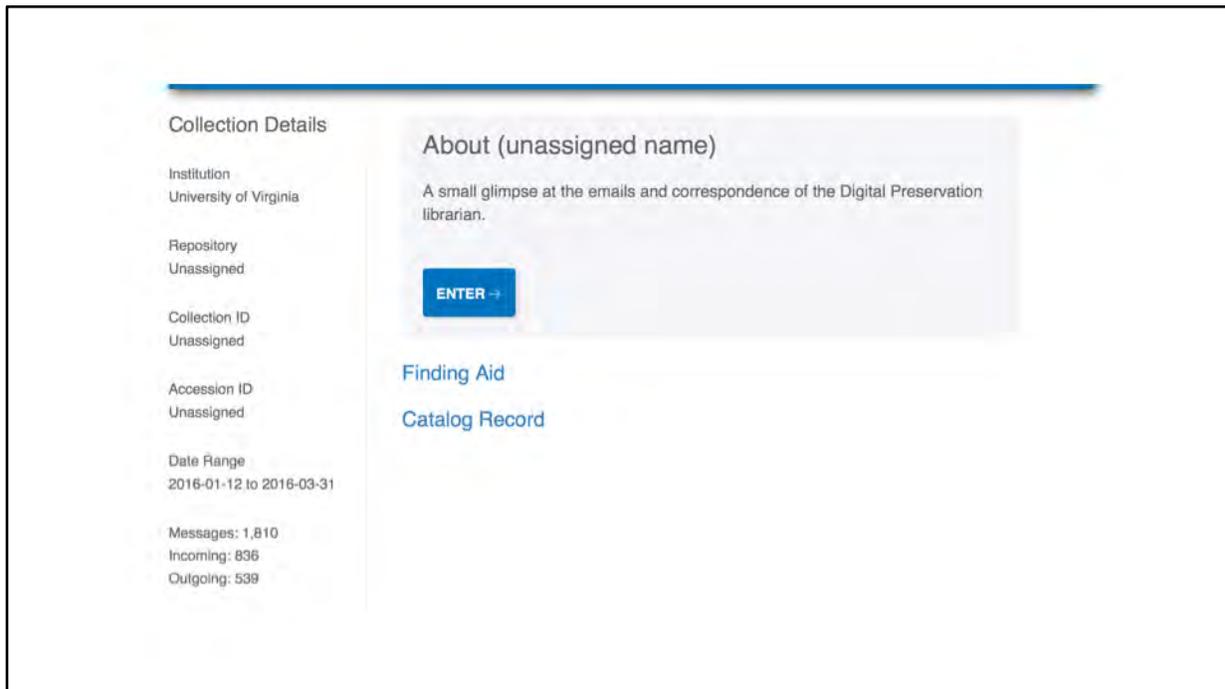
One tool we've been working with recently, and are currently developing a test donor use case for, is EPADD, which is an IMLS-funded project out of Stanford that is a software that support archival processes around the appraisal, ingest, processing, discovery and delivery of email archives.

So this tool is great in that it is modular in design, and the different modules (appraisal, process, discovery, etc.) can be set up in different part of the library. For instance, as the digital preservation librarian, I may be responsible for the initial ingest in the "appraisal" module., shown here, while Special Collections may house the discovery and delivery modules locally in a reading room.

This modular set up allows me to do things like review email for technical issues, help donors set up their initial ingest, review file types, and ensure content is safe and stable before moving the collection to another module, such as processing.

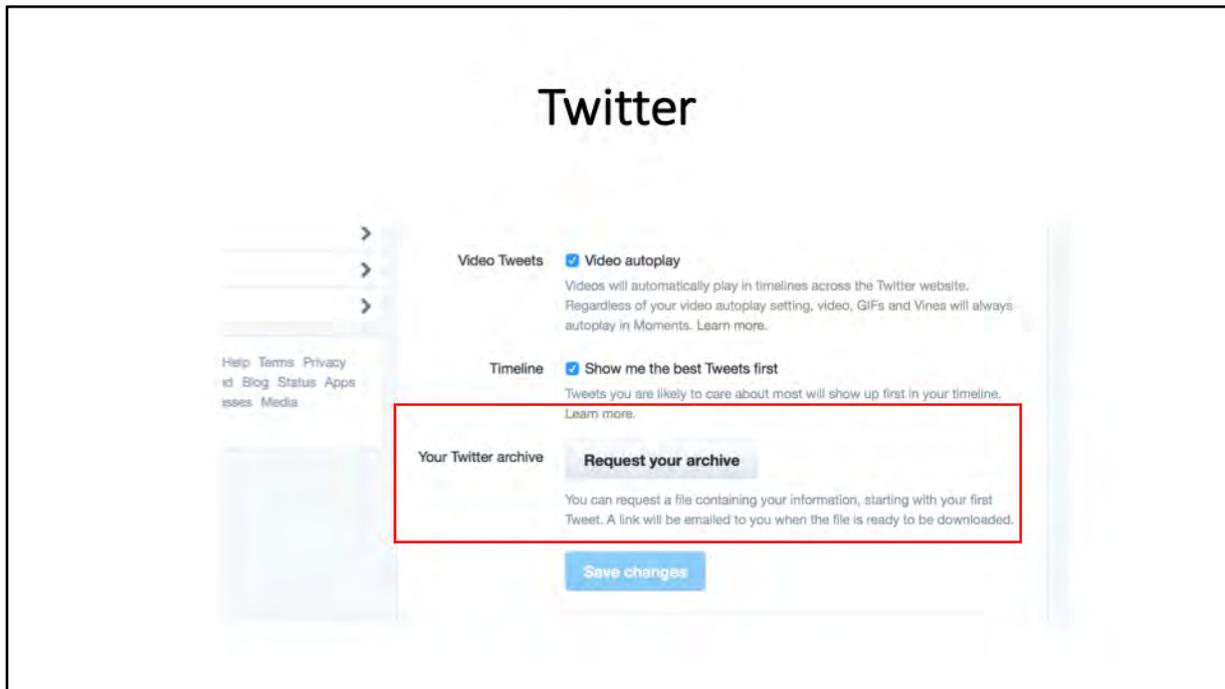
The amount of data related to contacts, dates, attachments, images, sensitive information, is extremely useful in this tool, and allows you to search, flag content, and view email (including chains of correspondence) to ready the collection for preservation and use.





This is a snapshot of what the processing module looks like in EPADD from a test I ran through my own email - you can see it's set up for use in Special Collections and archives with the ability to connect the email collection to both a finding aid and catalog record, the assignment with Collection and Accession IDs, and more.

We're currently in the testing stages of this tool and working with Special Collections on a use case, so I look forward to hopefully giving another talk on this preservation work in the future.



So what about social media? This is a newer digital preservation puzzle that also highlights issues related to privacy, scale, policy, use & access.

We have many active student organizations, departments and events that are related to the history and records of the university, not to mention donors who may be active on social media and desiring to deposit the records with the University. How can we preserve this material?

One way, especially when working with the personal archives of an individual, is, like email, to request their personal twitter archive for deposit - this give you the entire history of your tweets since the beginning of your account in stable, widely adopted formats that we can ingest into our repositories.

The screenshot shows a Google Sheet titled "Scalia TAGS test" with a menu bar (File, Edit, View, Insert, Format, Data, Tools, Add-ons, Help, TAGS) and a toolbar. The spreadsheet contains 10 rows of tweet data. The columns are labeled as follows: A: id\_str, B: from\_user, C: text, D: created\_at, E: time, F: geo\_coordinates, G: retweet\_count, H: retweeted, I: user\_lang, J: in\_reply\_to\_user\_id, K: in\_reply\_to\_tweet\_id.

A	B	C	D	E	F	G	H	I	J	K	
1	id_str	from_user	text	created_at	time	geo_coordinates	retweet_count	retweeted	user_lang	in_reply_to_user_id	in_reply_to_tweet_id
2			RT @OrMartyFox: Like A Good Little #Marxist #Obama Skips #NancyReagan & #Scalia Funerals								
712946	driscott_atlanta	Visits The Barbican: #Castro Brothers https://t...		Thu Mar 24 10:19:28	24/03/2016 10:19:28		574		en		
712946	BrianDa45124592	RT @TalksOnLaw: Will #Scalia be replaced by a recess appointment? @Judgenap explains! https://t.co/huAHK2qqk0 #SCOTUS https://t.co/bn6PKe0...		Thu Mar 24 10:19:04	24/03/2016 10:19:04		22		en		
712942	umakals	RT @TalksOnLaw: Will #Scalia be replaced by a recess appointment? @Judgenap explains! https://t.co/huAHK2qqk0 #SCOTUS https://t.co/bn6PKe0...		Thu Mar 24 10:03:47	24/03/2016 10:03:47		22		en		
712940	chnsaw5210	RT @TalksOnLaw: Will #Scalia be replaced by a recess appointment? @Judgenap explains! https://t.co/huAHK2qqk0 #SCOTUS https://t.co/bn6PKe0...		Thu Mar 24 09:53:06	24/03/2016 09:53:06		22		en		
712936	melianieusn1979	RT @CNN: Justice Clarence Thomas reads a Bible passage at #Scalia's funeral https://t.co/D9o45leW9o https://t.co/Z1t5eE1P		Thu Mar 24 09:36:23	24/03/2016 09:36:23		196		en		
712936	melianieusn1979	RT @CNN: Justice Clarence Thomas reads a Bible passage at #Scalia's funeral https://t.co/D9o45leW9o https://t.co/Z1t5eE1P		Thu Mar 24 09:36:23	24/03/2016 09:36:23		196		en		
712930	panibbi	@orizontescuola mAFINTELAsiete+ndicolci d #Renzi e #Scalia! Nn c son #recaABILITATI dserieAeBI #LUBUNITI lcaaxo xceti #recaABILITATI		Thu Mar 24 09:15:18	24/03/2016 09:15:18				it	70350451	ortiz
712904	Kimmyincall	RT @angelaczarville: Horowitz: Justice #Scalia's Final Dissent https://t.co/ShoAHST3TU He was the last of his kind... #RebootLiberty #ORRPLUV		Thu Mar 24 07:31:16	24/03/2016 07:31:16		31		en		
			RT @angelaczarville: Horowitz: Justice #Scalia's Final Dissent https://t.co/ShoAHST3TU He was the last of his kind... #RebootLiberty								

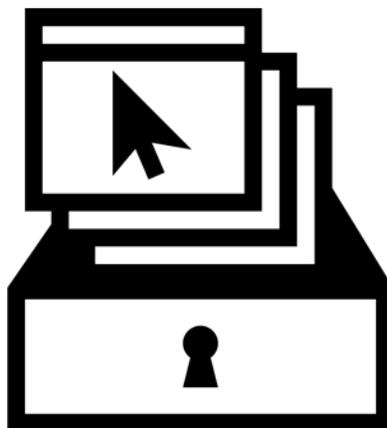
But sometimes we'd don't need to preserve the tweets of an individual - rather, we may want to preserve the records surrounding an event, such as the upcoming bicentennial at Uva. In this case, we have one tool currently in use called TAGS, which is open source and freely downloadable for anyone to use with the Twitter API.

This tool allows you to gather tweets around a hashtag in a simple Google spreadsheet format, which you can see here with a test use case I ran around the time of Antonin Scalia's death earlier this year. This tool is particularly useful in that it allows you to capture the immense amounts of metadata that are behind each 140-character tweet, which can be very powerful for preservation and future scholarly use.

We're also always following the digital preservation and archives community of practice and tool development. An example of a new project we're following is a Mellon grant-based project called Documenting the Now, which is aimed to address this event-based capture scenario of Twitter for scholarly use and preservation. DocNow is geared toward libraries and archives, who now find themselves stewarding and preserving content from social movements and historical events that largely take place online, such as the protests in Ferguson, Mo. in 2014, around which

the grant is based.

## Web Archiving

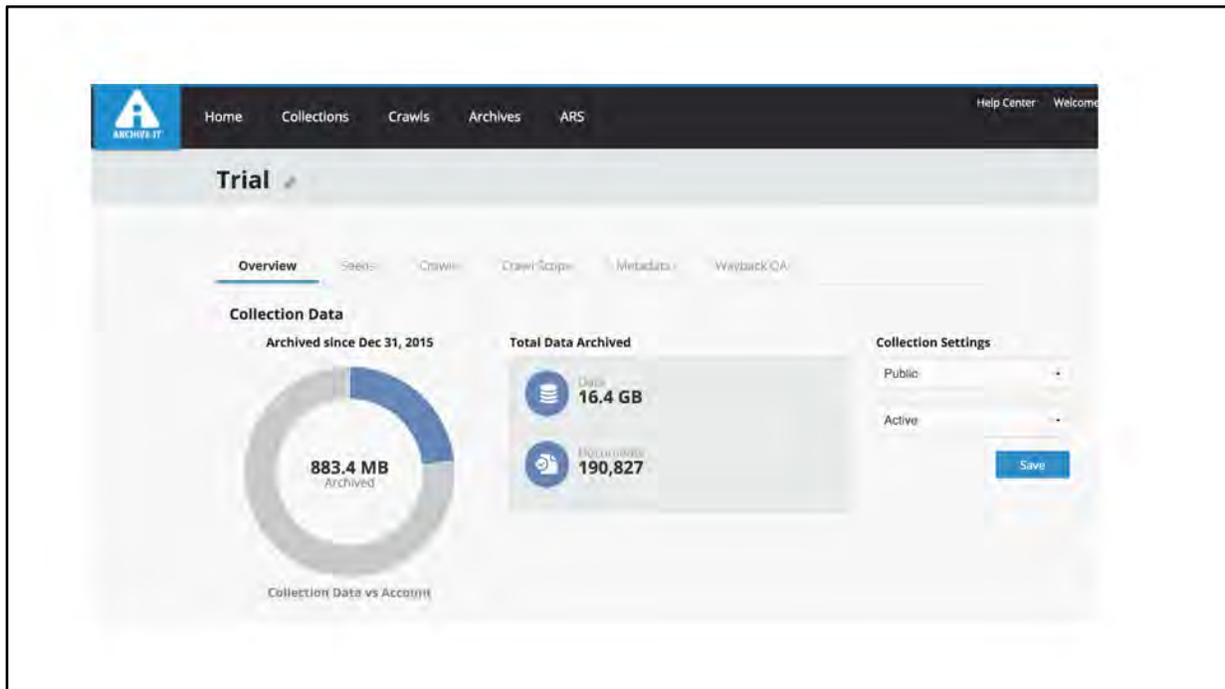


The last item I wanted to touch on today is how we're approaching web archiving at UVa as part of our digital preservation work.



And I thought I'd start by showing off the very first University of Virginia website page that is in the Internet Archive from 1997. This is our incoming Library Dean John Unsworth's page at the time.

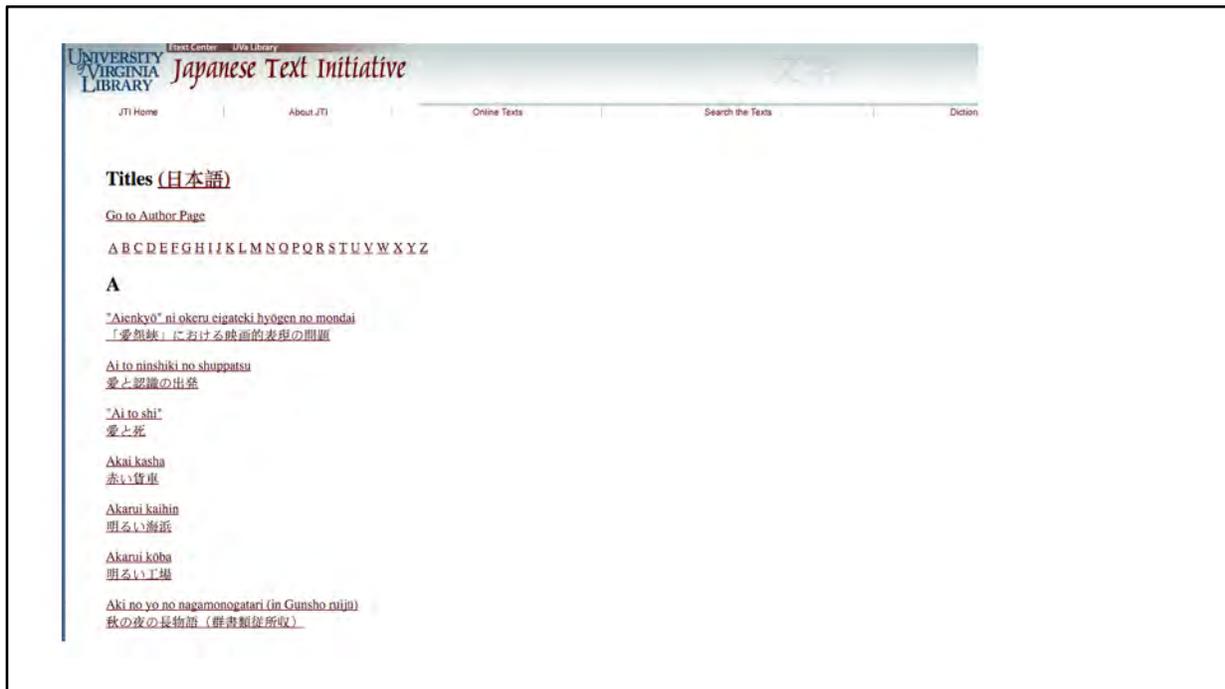
So if it's in the Internet Archive, it's preserved forever, right? Not so much. The wonderful, forward-thinking Internet Archive is many things, but they have explicitly stated that they are not, in fact a preservation repository nor do they plan to be. And I'm sure every single one of us, almost daily, stumbles into a 404 error, or goes to a website that suddenly seems to have been taken over only by ads. The internet is brutally ephemeral and ever-changing in its structure and content. So how are we approaching the preservation of web-based university content over time?



One of our approaches is through a service some of you may be familiar with called Archive-It. This is a subscription service from the Internet Archive that allows us to scope, test and run specific crawls on specific websites to preserve selected content at intervals we can set - this may be just a one time crawl, or an ongoing capture every two weeks.

These crawls are then available for download and preservation via Web Archive files, or WARC, which is an international standard for web accessible content in an archived state. These files contain both the primary resources in any format, as well as secondary content such as metadata, duplicate crawl events, and more.

One of the ongoing digital preservation projects we're working on related to web archives includes a consortial effort with 2 other AP Trust members - Penn State and the University of Miami - to create an AIT to APTrust preservation workflow that will allow us to send these WARC files directly for preservation in AP Trust.



So are we looking to preserve? Currently we're in the nascent stages of our program, but our focus through the testing stage over the last year or so has included capturing digital projects and scholarship such as the Japanese Text Initiative;

- ▶ CALENDAR
- ▶ UPCOMING BOARD MEETING
- ▶ UPCOMING COMMITTEE MEETINGS
- ▶ PAST MEETINGS
- ▶ PUBLIC MINUTES
- ▶ COMMITTEES
- ▶ COMMITTEE MINUTES
- ▶ BOARD MEMBERS/STAFF
- ▶ HOME

### Video Archive: Past Board of Visitors Meetings



Final Session 11/9/12



External Affairs Committee 10/19/12



Committee on Governance & Engagement 10/19/12



Full Board Meeting 10/19/12

University administration and life, including the Board of Visitors meeting minutes and streaming video, which are only online;

You are viewing a temporarily archived web page, collected at the request of University of Virginia using Archive-It. This page was captured in a test crawl on 18:36:00 Mar 29, 2016, run in the The Cavalier Daily collection. If you would like this crawl to become part of your collection permanently, you will need to save it from the [crawl report](#). The information on this web page may be out of date. See [All versions](#) of this archived page. [Videos](#)

Mar 29, 2016  
 News Sports Opinion A&E Life Focus HAS Humor Blogs Multimedia Brand Link

# The Cavalier Daily

Search

**Charlottesville Police join Facebook**  
 by Hannah Cho | 15 hours ago  

 The page is already reaching close to 1,400 likes after only four days of being live.  
[READ MORE >](#)

**Batten lecture explores global roles of China, Japan**  
 by Jack Tuttle | 15 hours ago  

 Much of Midland's lecture focused on the national security, economics and evolving identities of China and Japan and their relationship to the western hemisphere.  
[READ MORE >](#)

**Spring philanthropies abound on Grounds**

**U.Va. releases admission decisions for class of 2020**  
 by Catherine Wiedmann | 15 hours ago  

 The accepted students represent 102 foreign countries and 49 U.S. states.

**OPINION**  
**Trump chalking at Emory has been blown out of proportion**  
 by Editorial Board | 14 hours ago  
 Though the candidate himself may not be, the chalkings in question are relatively innocuous.

and student organizations and publications, such as the Cavalier Daily.

The image shows a screenshot of the Webrecorder Beta website. The main heading is "Welcome to Webrecorder Beta". Below it, the text reads: "An open source web archiving platform and service for all! Easily create high-fidelity, standards compliant archives of the web as you browse." Underneath, it says "Try now:" followed by a form with a text input field labeled "Enter a url here:", a "Record" button, and a "Preview before recording" checkbox. Below the form, it says "We are inviting users to build their own archives as we test this service. Request an invite to sign-up". It also mentions "A project of RHIZOME" and includes a "Donate" button with the text "to support Webrecorder and related open-source web archiving tools. Donations will be tax-deductible." To the right of the main content is a blue advertisement for "WARCreate Google Chrome Extension". The ad features a logo of a cardboard box with a green arrow pointing down. The text in the ad includes: "WARCreate Google Chrome Extension 'Create WARC files from any webpage'", "What is it?", "WARCreate is a Google Chrome extension that allows a user to create a Web ARChive (WARC) file from any browseable webpage. The resulting files can then be used with other tools like the Internet Archive's open source Wayback Machine. The tool is an evolving product with the end result pushing toward being a personal web archiving solution for those that wish to securely archive their metadata in a standardized way.", "Where Can I Download It?", "WARCreate can be downloaded from the Chrome Web Store.", "What Can I Do with Generated WARCs?", "Software is needed to replay the WARCs by nature of the format. I recommend using Web Archiving Integration Layer (WAIL), a software suite that came about because of WARCreate."

And that's not all we're doing. I mentioned the ephemeral and ever-changing nature of the web. Archive-It is the core of our approach, but as video, audio, and other information on websites continues to shift, grow and change, we also combine additional web preservation tools with our AIT service, such as webrecorder.io, a developing, open source software that more easily captures streaming content on websites, as well as tools such as WarCreate, a Google Chrome extension that enables anyone to download the Web Archive (WARC) files from any browseable webpage.

## Future Work

- Transfer of born digital content
- UVA workflow development
- Emulation

There's plenty of digital preservation work to keep me busy, but I wanted to touch on a few items that we're continuing to work on and work toward with digital preservation at UVA.

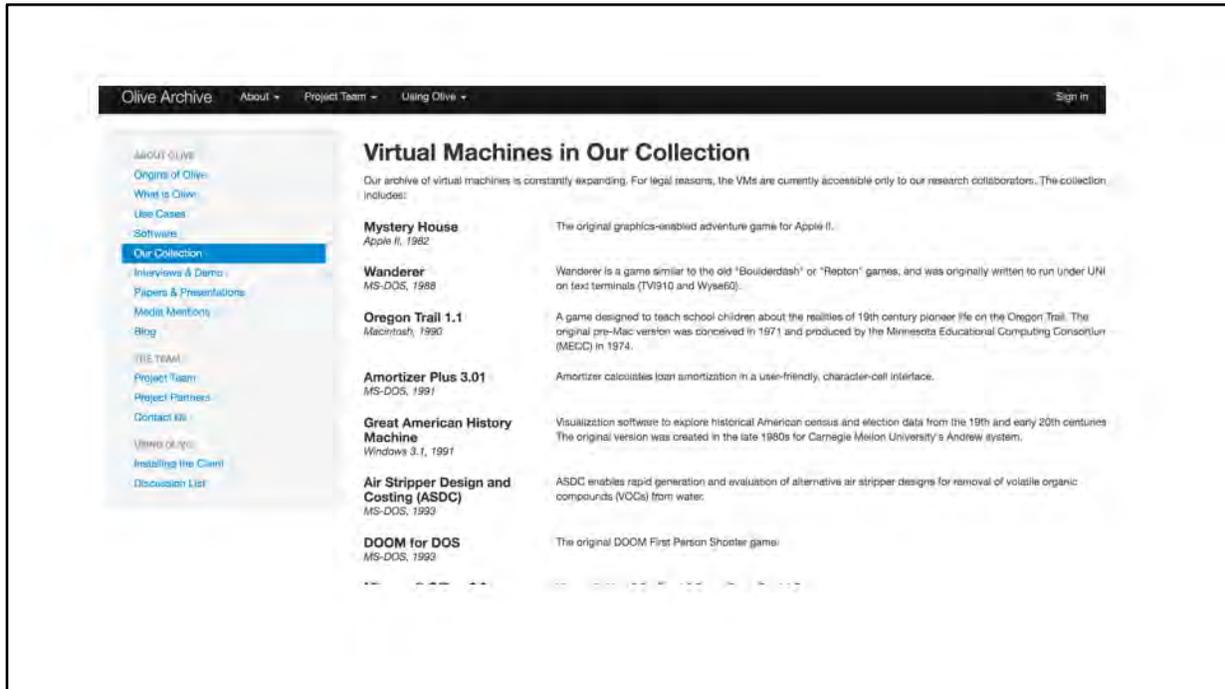
**The first** refers to the transfer of born-digital content. And by this I mean continuously working toward the best, most stable solutions for the transfer of born digital content to the library and archives for ingest, preservation and use - without needing a physical carrier such as a hard drive to do so.

There are developing tools and standards for born digital file transfer that allow for the preservation of file structure, allow us to provide metadata and create checksums and receive notification of delivery to allow the secure audit and transfer of important digital materials. So we continue to work to work toward that.

**The second**, related and very important ongoing work relates to connecting the workflows, procedures, and standards here at UVA. This is a large collaborative effort to connect my work with digital preservation to the workflows across the different departments in Special Collections, CAD, and others to ensure that we're working together to preserve, describe and make accessible the great works we have here at

UVa

The last is a bit further out for us as a process currently, but it warrants mentioning as part of striving to achieve the best standards for digital preservation. And that is emulation.



Emulation could provide a strategy to allow access to content by preserving executable content (such as databases, visualization software, etc.) in its original state and running it on virtual machines. This could replace or supplement other digital preservation strategies for file formats such as the migration of files to newer versions of the format.

The Olive project out of Carnegie Mellon is doing some very interesting work in this area, and we would be interested to pursue it here at Uva.

# Thank you!



Lauren Work  
Digital Preservation Librarian

University of Virginia Library  
Preservation Week 2016  
[lw2cd@virginia.edu](mailto:lw2cd@virginia.edu)



In closing, I hope you come away from today knowing a bit more about digital preservation as a whole, and a good deal more about all of our digital preservation work here at UVa.

Thanks.