

## Accessioning Digital Files: Tales from the Trenches



[www.bonanza.com](http://www.bonanza.com)

Nancy Enneking, Getty Research Institute

Lisa Miller, Hoover Institution Archives

Rebecca Wendt, California State Archives

Society of California Archivists  
April 2012

Welcome to Accessioning Digital Files: Tales from the Trenches

We chose “Trenches” to convey the idea of work in progress.

I've been thinking about this topic for several years, and I'm glad to see from the attendance that others are interested in it too.

## Archives 101 - Definitions

*Accession* - To take legal and physical custody of a group of records or other materials and to formally document their receipt.

*Ingest* - In the Open Archival Information System (OAIS) model, processes related to receiving information from an external source and preparing it for storage.



**A Glossary of Archival and Records Terminology**

RICHARD PEARCE-MOSES

Since we confronted accessioning our first file-based collections at Hoover I've asked many questions about accessioning, and consulted the SAA Glossary in search of answers.

- Exactly what is accessioning, and what are its outcomes?
- Can we accession digital files the same way that we accession paper materials?
- Does the scope of accessioning expand when digital files are involved?
- Does accessioning digital files equal ingest, or expand to include ingest?

## Archives 101 - Definitions

*Accession* (citing Ann Pederson, ed., Keeping Archives) - Having made sure that new material has been legally transferred to your archives, the next, and vitally important, step is to gain control over it. This initial process is called *accessioning* which records information about origins, creator, contents, format and extent in such a way that documents cannot become intermingled with other materials held by the archives. Accessioning provides the basic level of physical and intellectual control over incoming material....**Accessioning consists of a sequence of different activities. These include preliminary sorting of the accession, recording the essential identifying information about the material, and its creator in the accession register and providing suitable storage for the material.**



### A Glossary of Archival and Records Terminology

RICHARD PEARCE-MOSES

The Glossary's definition is pretty short, but this extra note describes an accessioning workflow and refers to an accession register, which really fleshes out the definition.

Specific to Hoover, I've wondered:

-Is everything that we need to do getting done when we receive digital files?

-How are other repositories--besides the major universities--accessioning this material?

The screenshot shows the website of the Society of American Archivists (SAA). The header includes the SAA logo and navigation links: Home, The Archives Profession, About Us, Education & Events, Publications, Members, and Groups. A search bar is also present. The main content area displays the title "Guidelines for Reappraisal and Deaccessioning - DRAFT" and a table of attachments. Below the table, there is a "Group Newsletter" section with links to newsletters from Winter 2012, Summer 2011, and Winter 2011. A large grey box contains a list of key points related to accessions.

| Attachment   | Size      |
|--|-----------|
| GuidelinesForReappraisalAndDeaccessioningDRAFT.pdf | 225.91 KB |

**Group Newsletter**

- [Acquisitions and Appraisal Section Newsletter, Winter 2012](#)
- [Acquisitions and Appraisal Section Newsletter, Summer 2011](#)
- [Acquisitions and Appraisal Section Newsletter, Winter 2011](#)

-Who owns the collection (importance of signed deed of gift)  
-Name of donor  
-Agreements, both formal and informal, with donor  
-Acquisition dates  
-Content and types of materials in the collection

As a result of thinking about accessioning, I'm on the lookout for evidence of its importance to our work.

One recent example is the "Guidelines for Reappraisal and Deaccessioning," which are undergoing review as an SAA Standard. To reappraise and deaccession you need to know a lot of information that is documented during the accessioning process.

This is just one of many archival activities that shows how critical accessioning is.

Why was it hard to find speakers for this session?

- I don't know enough people
- I don't know the right people
- People are not accessioning digital files
- People are not comfortable talking about how they're accessioning this stuff
- All of the above

Wanting to know what others were doing, I decided to create this session. But then I had trouble finding people to talk about this topic. I asked at least a half dozen people who declined. There are a lot of reasons for this, but

CLICK AGAIN

the important one is number 4, which means for us today that

- (1) Nancy and Rebecca get a lot of credit for being willing to talk about what their organizations are doing.
- (2) We must remember that this is a work in progress for all of us—hence the trenches or shower curtains--and none of us claim to be experts or models to be followed.

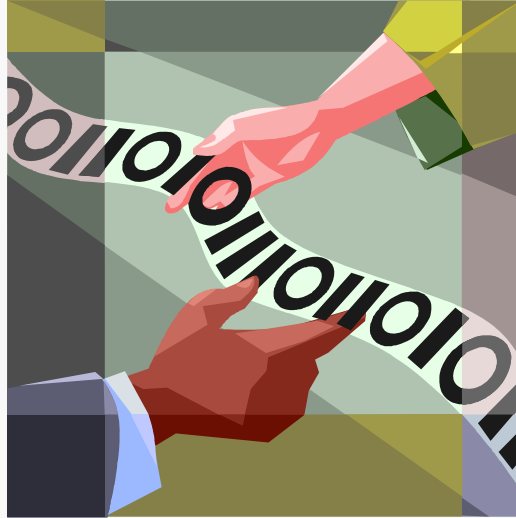
Nancy Enneking is the Head of Institutional Records and Archives in the Getty Research Institute at the J. Paul Getty Trust. Prior to her arrival at Getty in 2004, she spent 5 years with the Texas State Archives and previously worked at the Center for American History (University of Texas at Austin) and in the institutional archives at Johns Hopkins University. Nancy has a BA in History, a MA in Egyptology from Johns Hopkins University, and a MLIS in Archives and Records Management from the University of Texas at Austin.

Rebecca has been an archivist at the California State Archives since 2001. Like most of her colleagues, she wears many hats. Currently she wears the hat of both Electronic Records Archivist and Legislative Records Archivist (among other things). Prior to landing at the State Archives she was a Manuscripts & Photographs Curator at UC Davis Special Collections. And, even prior to her UC Davis experience, she was the Archivist and Records Center Coordinator for Yolo County. She has pursued graduate-level study in Public History, CSU Sacramento and received her MLIS from San Jose State

## Accessioning Digital Files at Hoover: Version 1.2

Lisa Miller  
Hoover Institution Archives

Society of California Archivists  
April 2012



I'll be talking about how we're accessioning manuscript collections at Hoover.

We're affiliated with Stanford but are largely independent of it.

That means we have a different staffing structure and workflow than Stanford Libraries. So we've created different accessioning workflows for digital files than what the Stanford archivists probably talked about this morning.

I'll be talking about our process at Hoover from more of an administrative level.

***Born-Digital Materials***

- 3,000+ audio files from Commonwealth Club
- 500 photographs taken by Giles Udy
- 335 transcripts of Uncommon Knowledge TV/webcast programs
- 35 text files from Ginetta Sagan

***Digitized Materials***

- 10,000,000+ page images from Iraq Memory Foundation
- 90,000+ page images of Estonian KGB files
- 500 photographs from Edward Teller family
- 100 page images of Chen Jiaxin diaries

Since 2008 we've accessioned more than a dozen new collections or accretions of file-based collections.

Some were born digital. Others were digitized by the creators or other parties, but we only acquired the digital copies.

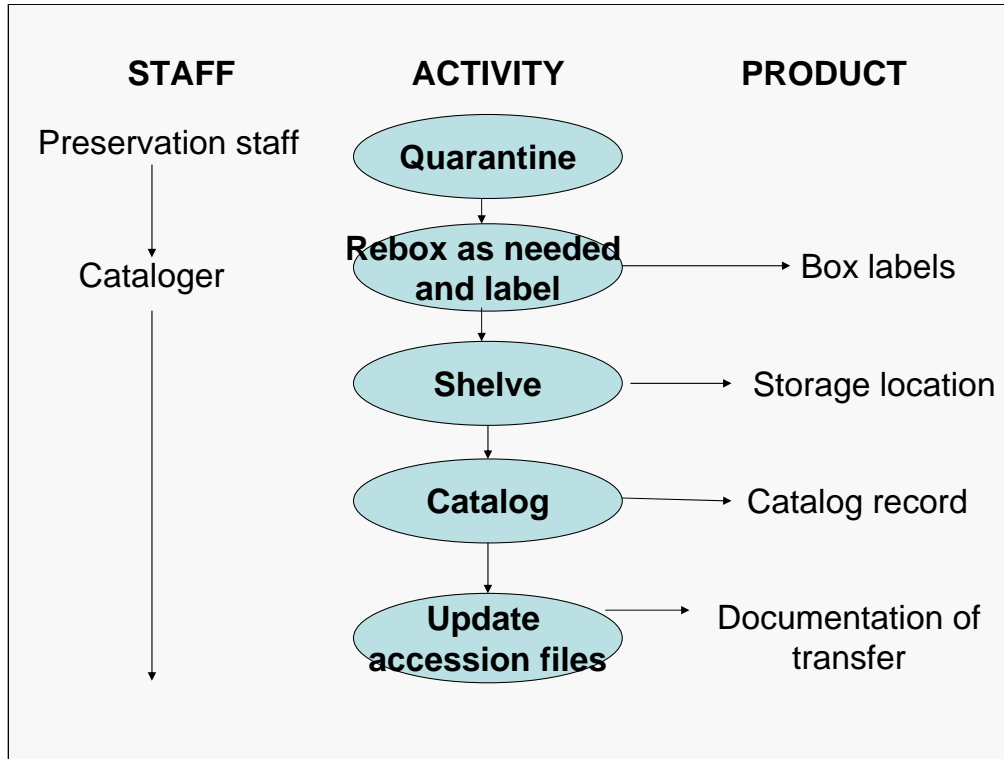
- Portable hard drives
- USB flash drives
- Email
- Downloaded from Internet



At Hoover, digital files that come in on CDs or floppy disks are boxed and shelved during accessioning, just like paper materials. So they are handled according to our standard accessioning workflow. Because digital files can't wait years for processing like paper materials can, we try to prioritize them for attention.

The collections I'll be talking about today came in over a network or on carriers that serve only for transport. They are reusable carriers that lack any special labelling data provided by the creator.





Our first file-based acquisition was an accretion to an existing collection that came on a portable hard drive in 2008. One of our tech guys brought in the hard drive, copied the contents onto our preservation server (we do not have a digital repository system), and considered the job done.

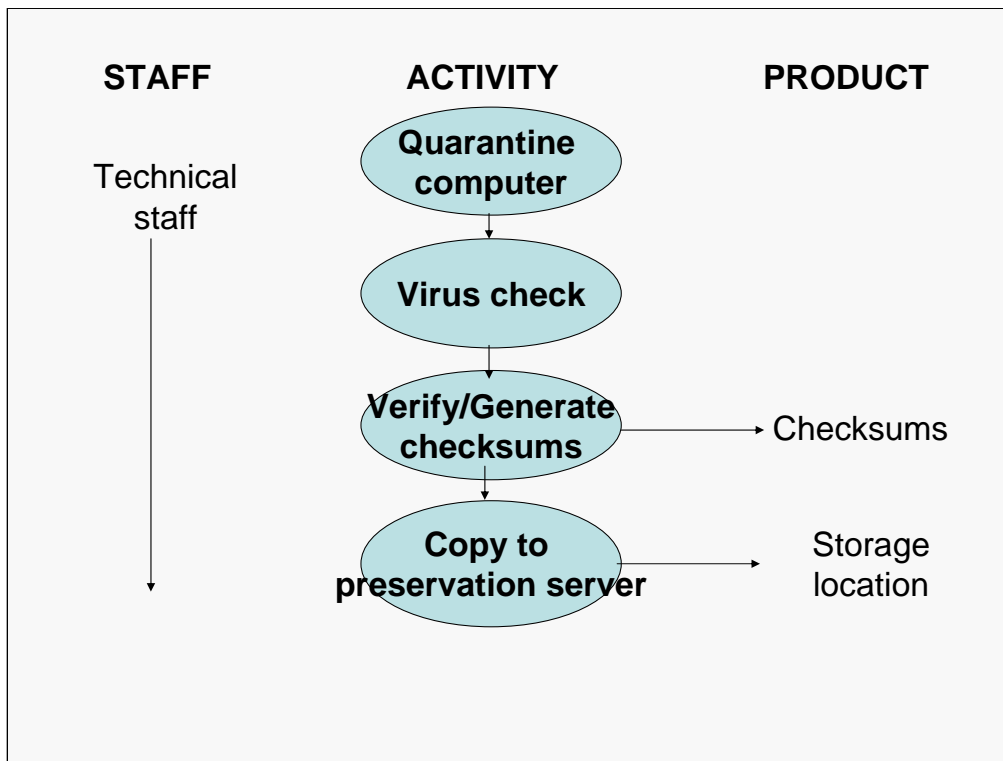
The tech people I'm talking about are part of the archives staff and do archival work, not IT work, but most of their training is in more specialized areas like AV, so they don't always fully appreciate archival basics.

So this tech guy did not realize that no one else in the archives knew we had the materials. The curator couldn't acknowledge the donor, the catalog record and finding aid were not updated, and reference staff and researchers didn't know about the materials.

So the entire accessioning process had been bypassed. We obviously needed a new workflow for file-based materials.

So I thought about our accessioning workflow for paper materials, and the products of accessioning. They centered on the cataloger.

He takes control of the material, provides storage and shelf location, and records essential data about the material and its creator in the catalog record—which is our accession log.



So then I thought about the way digital materials come in and get processed.

It is all technical steps that don't involve the cataloger. It doesn't generate the vital accessioning products that we log for paper materials, like a catalog record and transfer documentation.

The cataloger was key. Only he can add or update records in our catalog.

But he was not going to perform these technical accessioning steps, nor examine the files to glean basic data about creator, content, format, and extent.

So our tech staff needed to relay this data to the cataloger to get the materials officially accessioned. This was both a communication problem and a documentation problem.

**DIGITAL**

**ACQUISITION [or] INCREMENT [choose]**

**Date:**

**To:** Accession file, preservation server,  
[cataloger, other staff],

**From:**

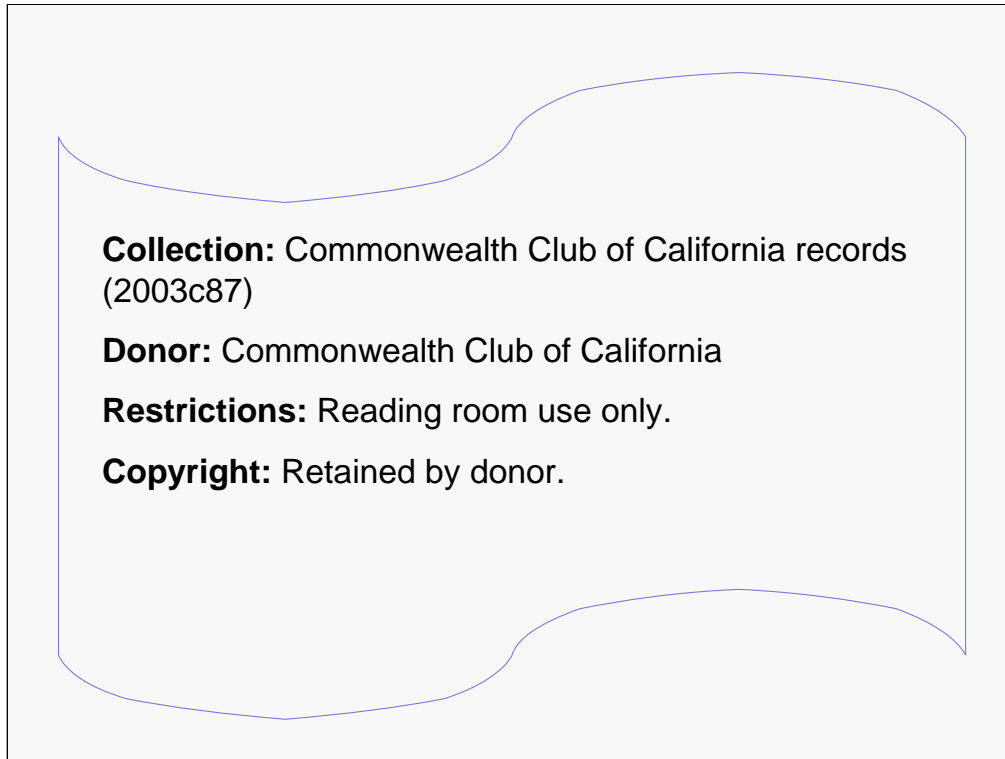
Maybe it shows my 15 years as a government employee, but my solution was to create a form. Our tech staff would fill it out as completely as they could, getting input from the curators or other staff.

Then they would share it with everyone who needed to know about the collection. I'll show all the elements of the form on the next few slides.

These elements contain some of the basic administrative information:

- Whether it's a new acquisition or accretion
- The date, which serves as a rough date of receipt and basic processing of the materials
- Who was responsible for the accessioning work

Not only is the form sent to staff, but a paper copy gets printed and added to the accession file, and an electronic copy sits with the files on our preservation server.



**Collection:** Commonwealth Club of California records  
(2003c87)

**Donor:** Commonwealth Club of California

**Restrictions:** Reading room use only.

**Copyright:** Retained by donor.

These elements form part of the catalog record.

We need this data whether the accession is paper or digital. We can usually get it from the curator or deed of gift, but it makes sense to get it all in one place on this form.

One workflow problem is that the cataloger officially assigns the collection title and its accession number. It's not a problem for accretions, which already have a collection title and number, but for some new collections the form is sent to the cataloger as a draft.

After the collection title and number is assigned by the cataloger, the tech staff finalize the form. They also have to wait to load the files onto our preservation server because we store our digital collections by accession number on the server.

**Series title(s):** Sound recordings of Commonwealth Club programs

**Series dates:** 2003 May 13 - 2008 January 26

**Series description:** Recordings of speeches and other events presented by the club in San Francisco and its regional chapters in the San Francisco Bay Area.

**Language of materials:** English

These are the archival description elements. There's nothing special or technical about them except that our tech staff, rather than the cataloger, describes the materials. This fleshes out the abstract in the catalog record, and can be plugged into a finding aid.

It's not meant to be the last word in description, but just an overview of the major series or types of materials if they can be identified. The form prompts for series because that provides more information for downstream use by our staff, and in many cases the material we receive is a discrete series.

If the collection is a hodge-podge of files—everything on a creator's laptop—we would complete this very generally rather than indicating series.

**Extent:** 1,685 programs (1,685 files,  
360,150,339,892 bytes or circa 360 GB)

**File specifications:** AIFF file format, 16 bit, 44.1  
ksp/s

**Digital creation dates (for digitized copies):**  
N/A

**Hoover storage location:** Preservation server  
(Weegie), folder 2003c87

So far all of the elements are basic to paper or digital accessions, and the unique thing is that our tech people are compiling the data. The rest of the elements focus data specific to digital files.

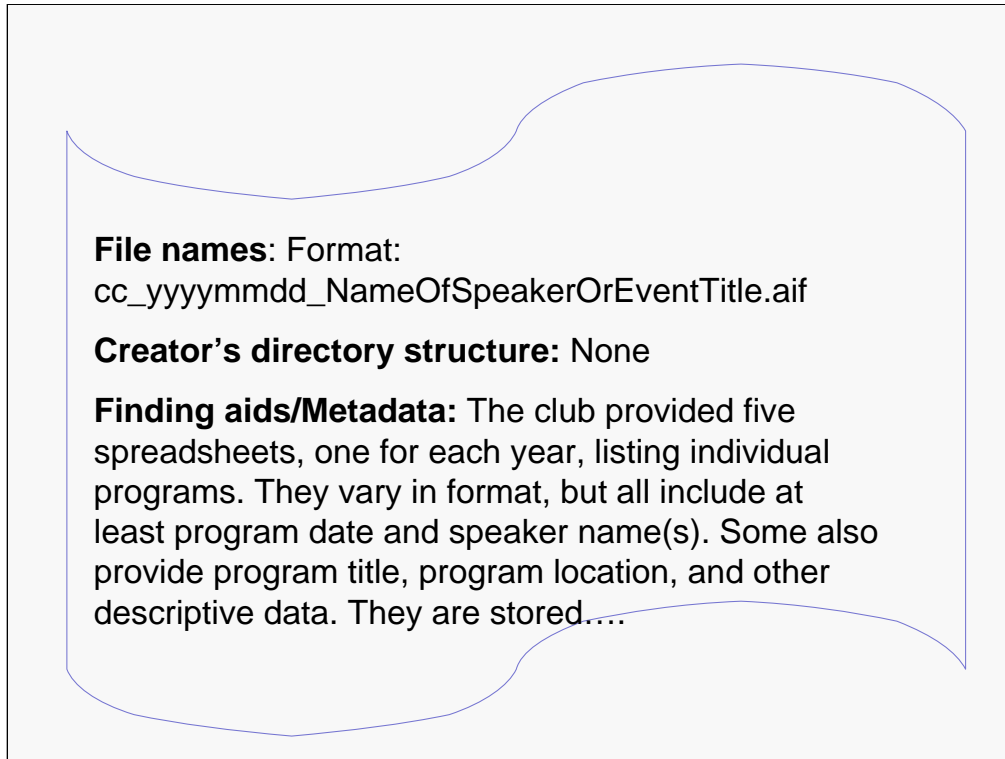
These elements capture basic technical details and are pretty quick to generate.

-For extent, we try to provide several measures. One is for the intellectual content (1685 programs). Another is for the number of files and their size. The exact byte size can informally verify the transfer, and the rough size helps us monitor space on the server.

-File specs show whether we received high-quality master files or low-quality use copies. Often it's only low-quality files because that's how the creator digitized their materials before we got involved. We don't generate and record technical metadata for every single file, so this form documents it once for the whole accession.

-Digital creation dates might be useful for managing the files later on.

-Most of our digital masters are stored on our preservation server, which we call Weegie, but sometimes we store file on other servers. Weegie was the name of Herbert Hoover's dog.



These elements help understand the content and arrangement of the digital files in a folder structure.

With paper materials we might ask about finding aids or file plans.

With digital files, the descriptive information is often embedded in file names or the creator's directory structure, so it makes sense to record data about them.

**Checksums (creator, date, software used):** MD5 checksums created by Hoover on 1 Jan 2009, using Fastsum, for content files on portable hard drive. Checksums created before files were copied from the creator's portable hard drive.

**Notes:** These are born-digital files that were transported to Hoover on the creator's portable hard drive. These files are preservation masters; derivatives must be created by Hoover as use copies.

Checksums are the digital fingerprints that uniquely identify each file. We use them to check file integrity after copying files, and to periodically check everything on our preservation server. If a file changes, it will not match its fingerprint, and we know that we need to take action.

The accession form reports who first created the checksums, and when. This is important provenance data. It indicates the exact point from which we can confirm that the files are identical to what we originally received. The earlier in the process the checksums are created, the better.

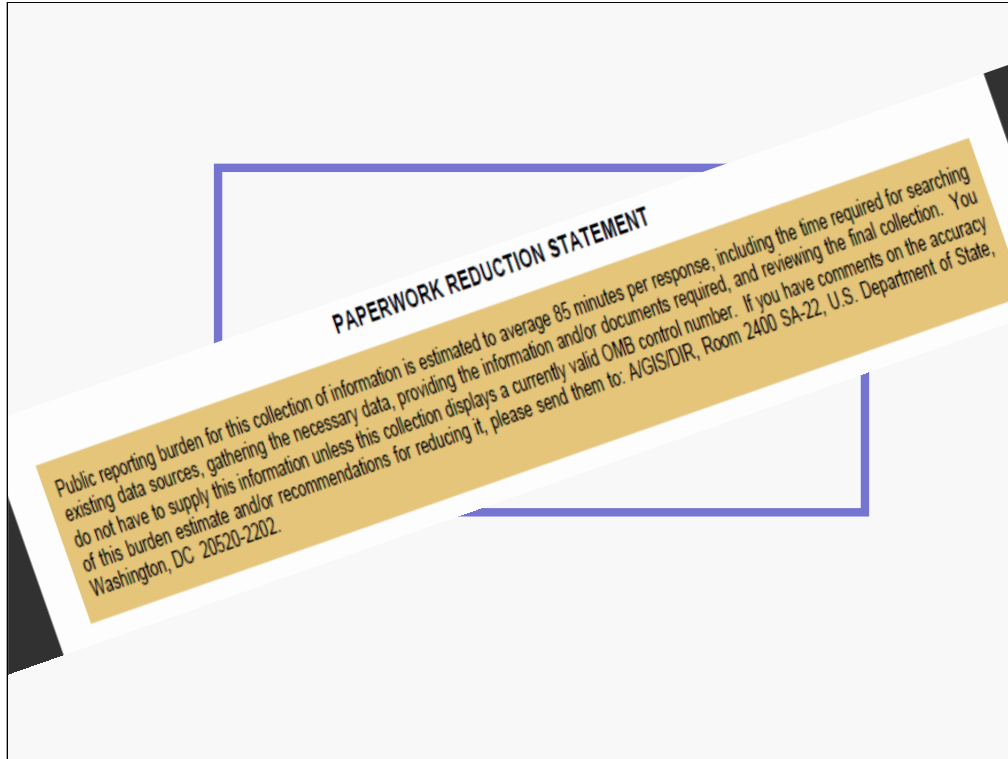
Just about anything can be added to the Notes field, including:

- details about how the materials were transferred to us—by flash drive or email.
- processing steps during accessioning, like whether we changed any file names.
- issues noticed during accessioning, such as image quality problems
- a directory tree for the files (which is generated by DOS commands)

This form gets the details of every digital accession into Hoover's institutional memory—by being added to a catalog record, printed and filed with the accession files, and residing electronically on our preservation server with the digital collection.

It's amazing how quickly one forgets the details of an accession. We've already consulted them as we maintain and manage our digital collections.





This is a lot of information to collect on incoming digital files, and it can be kind of tedious to write it all down.

How long it takes varies with the size and complexity of the collection, and I haven't calculated the average number of minutes to complete the form.

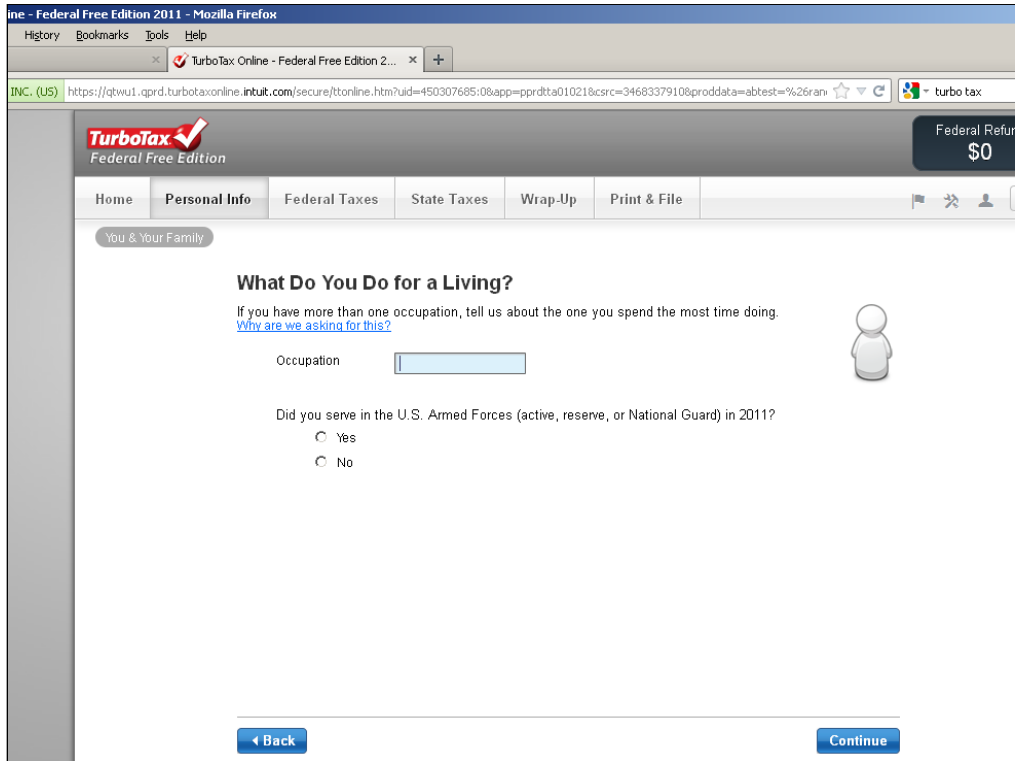
Fill out digital accession form

Get colonoscopy

I thought our tech people--who are archives people--would use the form if it was provided.

CLICK AGAIN

But it turned out some of them would rather get a colonoscopy than fill it out. So after introducing the form I realized that I needed to try to win them over to it.



It would help if the form was more interactive, kind of like Turbo Tax, walking you through the data elements with prompts, and measuring and reporting your progress toward completion. But we don't have programmers to do that.

## Educate on the importance of accessioning



File:Moofushi Kandu fish.jpg by Bruno de Giusti -- [http://commons.wikimedia.org/wiki/File:Moofushi\\_Kandu\\_fish.jpg](http://commons.wikimedia.org/wiki/File:Moofushi_Kandu_fish.jpg)

So I tried to educate them on the importance of accessioning.

One analogy I used was if they acquired paper materials, put them in a box with a label, and placed it in the stacks. The papers would be preserved, but no one would ever know they were there—one undocumented box lost in a sea of properly accessioned boxes.

They seemed to get this, but it didn't lead to any completed accession forms.

The image shows a screenshot of a web browser displaying a Wikipedia article. The browser's address bar shows the URL "wikipedia.org/wiki/Nagging". The page title is "Nagging". Below the title, there is a brief definition: "From Wikipedia, the free encyclopedia" and "Nagging is a form of pestering, or otherwise reminding an individual of previously discussed dictates or advice, usually from a perspective of superiority. The Scandinavian *nagga*, which means "to gnaw".<sup>[1]</sup>". A table of contents is visible on the left side of the article, listing sections: 1 Social nagging, 2 Parental and child nagging, 3 See also, 4 References, and 5 Further reading. A large, semi-transparent watermark with the word "Nag" in a bold, sans-serif font is overlaid on the bottom half of the page.

So then I nagged them.

**Amount of Material:**

**Series title and dates:** Sound recordings of club programs

**Series dates:**

**Series description:** Recordings of speeches and other events presented by the Club in San Francisco and its regional chapters in the San Francisco Bay Area.

**Language of materials:** English

**Hoover storage location:**

**File specifications:**

**File names:**

**Finding aids/Metadata:**

**Restrictions:** Materials are open.

**Copyrights:** Retained by the club per original agreement.

**Notes:** These are born-digital files that were transported to Hoover on portable hard drives. These files are designated as preservation masters by Hoover; smaller derivative files must be created by Hoover as use copies.

Partially complete form to get them started

Then I tried filling out as much of the accession form as I could and sending it to them as a way to help get them started.



Then I nagged them some more.

Cite in staff performance appraisals

HOOVER INSTITUTION  
STAFF JOB PERFORMANCE EVALUATION

- Quality of work:* Extent to which completed work is accurate, thorough, well-organized, neat.
  - Efficiency:* Consistency of performance, ability to plan and organize work, to set priorities, complete tasks promptly, and establish realistic goals.
  - Skills:* Adequacy of training and skills (including speaking and writing) to requirements and needs of the position, level of expertise in subject areas.
  - Attitude:* Sense of responsibility for work, responsiveness to suggestion and direction, job interest, attendance.
  - Relations with others:* Ability to cooperate with co-workers, respond helpfully to needs of scholars, faculty, staff, and communicate effectively.
  - Independent performance:* Degree of self-reliance and initiative in performing assigned work, correcting errors, identifying and solving problems, assuming new or enlarged tasks, reliability, and judgment.
- NOTE:* These following categories apply only to staff members with administrative or supervisory responsibilities.
- Administrative knowledge:* Understanding of policies, interdepartmental relationships, departmental routines, and degree of application.
  - Effectiveness:* Effectiveness in analyzing problems and needs within area(s) of responsibility, and in developing policies, procedures, and new programs.

I tried mentioning it on their performance evaluations.



# NAG



File:Nags Head town welcome.png by Sinneed -- [http://en.wikipedia.org/wiki/File:Nags\\_Head\\_town\\_welcome.png](http://en.wikipedia.org/wiki/File:Nags_Head_town_welcome.png)

And then I nagged them some more, and continue to do so.

It has gotten better—we've gotten a few forms completed—but we still have several accretions that have been in our custody for several years but remain unaccessioned.



File:Donuts.jpg by Lucianvenutian -- <http://en.wikipedia.org/wiki/File:Donuts.jpg>

Now, there are some things I haven't tried.

I didn't try food.



<http://www.pbs.org/wnet/nature/episodes/the-wolf-that-changed-america/wolf-wars-americas-campaign-to-eradicate-the-wolf/4312/>

And I haven't tried some sort of bounty system for completed forms.



Instead, our next step will be to create a storage space that we're calling Purgatory.

It will be a rest stop on the way to the preservation server. Tech staff will only be able to upload files to Purgatory, where we will review the files to make sure they are accompanied by an accession form. Only after this is confirmed will we move them to our preservation server.

Granted this is somewhat cosmetic—the files will be safe and backed up while in purgatory—but our tech staff is very committed to the concept of our preservation server, so I'm thinking that this will finally ensure that all digital accessions are truly accessioned.

If it does work, I might decide that Hoover's reached version 2.0 in its digital accessioning systems.

## **Accessioning Digital Files at Hoover: Version 1.2**

Lisa Miller  
Hoover Institution Archives  
lisa.miller@stanford.edu



I've been a bit irreverent in telling this story, but the bottom line is that accessioning is a vital archival process and we need to think about how it will get done when we acquire file-based collections.

I look forward to hearing how my colleagues are doing it at their repositories.