Three fundamental questions about personal digital archiving

What's in a personal (digital) collection?

Where should we store personal collections?

What can we expect as far as personal stewardship?
Three obvious answers

- Keep everything
- Teach archival best practices
- Develop a safe central repository with sophisticated access policies

- Digital stuff
- Human activity
- Technology
What would happen if we subjected these three obvious answers to the Snopes test?

Ratings Key

- = true
- = false
- = multiple truth values
- = undetermined
- = unclassifiable veracity
Rumor #1

Q: What’s in a personal digital collection?

A: Everything. Storage is cheap; keep it all!
Storage *is* cheap:

*a terabyte personal store*
So what are some of the assumptions behind the directive to keep everything?

- It’s an unalloyed good thing to be able to recover everything you’ve ever created or encountered.
- Deletion is hard, thankless work.
- We can use methods of filtering and searching to locate the gems among the gravel.
Is keeping everything a good idea? It *is* from a memory prosthesis view...
Furthermore, deletion is hard work...

“[when I buy a new computer] I transfer everything. ... [The computer] is the same [except] it’s faster. I should take the time to clean it up at that point, but [I don’t].”

When asked when he ever got rid of digital stuff, one person I interviewed said,

“Yes, but not in any systematic manner. ... It’s more like, I have things littering the desktop and at some point it becomes un navigable...

A bunch of [the files] would get tossed out. A bunch of them would get put in some semblance of order on the hard drive. And some of them would go to various miscellaneous nooks and corners, never to be seen again.”
And people use loss as a means of deaccession

“If [my email] were totally lost it wouldn’t be the end of the world. I guess that I don’t consider anything tangible, like, so important as an emotion or an experience, I guess I’m kinda of like a Buddhist.”

“If my hard drive was gone, it really wouldn’t bother me all that much, because it’s not something I need, need. I just thought it would be nice to keep it around.”

“I mean, if we would’ve had a fire, you just move on.”[re: 13,000 email messages that participant has saved intentionally] “And they’re all stored in here. On the computer... Never have [backed them up]”

[from researcher interviews] “Unfortunately I use a lot of data that is very very big, gigabytes of stuff... and it's not backed up. It's a bad situation. But what can you do?”
In other words...

It’s easier to *keep* than to *cull*...
But we also know that

- although storage is cheap, human attention is less so.

- it’s not legally or emotionally viable to keep everything.
Is it any different for personal scholarly archives?

- much of what we keep today does not seem to have long term value: personal scholarly archives are not just the file system + 20 years.

- what do researchers value?
  - PDFs of publications
  - Some bibliographic resources

- what might they value?
  - Datasets (in some domains they clearly do)

- whatever we keep must be disentangled from the institutional storage
Storage is cheap; keep it all!

Ratings Key

- = true
- = false
- = multiple truth values
- = undetermined
- = unclassifiable veracity
So it seems like what we’re looking for is the ability to expunge PLUS the digital equivalent to benign neglect PLUS the ability to safeguard the things we really care about...
Value-related strategies for keeping digital stuff

known high-value stuff

medium value stuff—we used to rely on benign neglect to leave us w/some of it

Preservation through use—*the more I use these items, the more valuable they’re likely to be and the better their chance of survival*

lower value stuff—ambivalent attitude

the controversial stuff—deleted? Are you sure? Forensically?
Q: How can we foster personal stewardship?
A: Teach archival best practices!
Saving files with a CD-RW drive

Q: My Windows XP computer came with a CD-RW drive but not a floppy drive. Is there a way to use files on a CD-RW drive as if it were a floppy drive? Do I need to buy software for this or is it a simple process?
A: It's not as simple as it seems. You would need a steep learning curve for a relatively straightforward process. You would not require a relatively straightforward process. However, you can use software that can format a CD-RW drive as a floppy drive. This software can be found online or in software stores. It's important to note that not all CD-RW drives are compatible with all operating systems.

David Einstein
Computing Q&A

Q: I recently switched from Internet Explorer to Firefox. With a free browser like Firefox, I could easily check my emails without a separate email service provider. But when I switched, I couldn't use Outlook Express any longer. How can I still use a spell-checker with Outlook Express now?
A: Here's the deal. AOL has its own built-in spell-checker. But if you don't have it installed, you won't be able to use a spell-checker. First, you need to download the free spell-checker for Outlook Express from Microsoft's website. Then, you can install it on your computer. This will give you access to a spell-checker for Outlook Express.
Some people aren’t even sure they believe in digital stewardship...

It’s funny though. If you look at technology, it’s just one of those things. I mean, whose fault is it? Is it the user’s fault for not backing up? Or is it technology’s fault for not being more tolerant and failsafe? In ten years, maybe hard drives and PCs will be so invincible and the Internet will be so pervasive that the concept of backing up will be quaint.

participant in an interview study who had lost his personal and business websites in a crash
Teach archival best practices to whom?

The home archivist is not always the home IT person
“I tried to install it [Firefox] and then John [her ex-husband] said, ‘Don’t install anything on your computer.’... I usually defer to John. Because he’s the one that’s got to come over and maintain it. So I have to make sure that it’s okay with him. But Jack [her 18 year old son], y’know, Jack will just do whatever he wants.”

“The conundrum that I’m in is [that] in order to back anything up on this computer, the computer has to be working well, and in order to get the computer working well, I should have backed up everything on this computer.”

“It’s kind of weird but with some of these CDs you can tell how much is written on it by looking.”
People don’t just rely on technology... digital stewardship is a social activity...

“Even my personal statement was saved onto that computer [the virus-infected laptop]. Then luckily, I also emailed it to my cousin, Camilla, at her house. ... So I said, “Camilla, do you still have my UCLA personal statement. She’s like, “Yeah.” So I said, “Okay, can you please email it.” So then that’s how I actually got it back to this computer.”
And remember the scale: there are over 3.2 billion personal photos on Flickr (this one is #2,975,981,038)

and that’s just **Flickr**: Facebook has over twice that many!
people may well be getting
better at capturing
better at creating
better at shaping
better at sharing
"i hosted my podcasts early on on a free service called Rizzn.net... he then changed rizzn.net to something called blipmedia.com and then!! he decided to sell blipmedia ... and he never emailed people about it.. suddenly the files were gone and the only news i heard about it was when i had to hunt online for what happened..."
It’s easier to *keep* than to *cull*,

but it’s easier to *lose* than *maintain*.
Teach archival best practices

Ratings Key

- = true
- = false
- = multiple truth values
- = undetermined
- = unclassifiable veracity
curation services and mechanisms

• invisible routinized activities that can be automated as services
  – E.g. find files that need any form of canonicalization at deposit (some people save important photos in RAW format)

• communal organizing and labeling activities (harnessing the power of social networks)
  – on an individual level, tags, annotation
  – on an institutional level, format registries

• everything else (stuff requiring human intervention)
Rumor #3

Q: Where should we store this stuff?

A: develop a safe central repository with sophisticated access policies
[11:09:24 PM] g says: [There are] 6 [online places where I store things] in all. 1.) school website, 2.) blogspot, 3.) wordpress.com (free blog host, different from wordpress.org), 4.) flickr, 5.) zooomr (for pictures, they offer free "pro" accounts for bloggers, but even for non-pros, they don't limit you to showing your most recent 200 pics only unlike flickr), 6.) archive.org

[11:10:42 PM] Cathy Marshall says: I ask just because you seem to have stuff in a lot of different places (so far two different blog sites, flickr, youtube, msnspaces, ... maybe yahoo?)...

[11:11:07 PM] g says: oh right.. youtube because people always tell me that they don't feel like downloading my quicktime files from archive.org

so people put copies of their stuff in different places for different reasons.

and safety is an important side effect!
develop a safe central repository with sophisticated access policies
It’s likely that there’s no single repository!

- Catalog driven approach
- Personal archives probably involve multiple storage sites
  - e.g. free software to create S3 backup
  - e.g. for-pay “vault” software
- Use different methods for handling medium- and low-value items
- All solutions should acknowledge the human tendency toward benign neglect
Summing it up

- Storage is cheap; keep it all!
- Teach archival best practices!
- Develop a safe central repository with sophisticated access policies!

Ratings Key

- Green dot = true
- Red dot = false
- Multiple dots = multiple truth values
credits

• personal digital archiving field study collaborators: Sara Bly and Francoise Brun-Cottan

• Web site recovery study collaborators: Michael Nelson and Frank McCown (ODU)

• Catharine van Ingen, the Community Information Management project at MSR SVC (Doug Terry, Ted Wobber, Tom Roddehoffer, Rama, and Rama Kotla)
any questions?

contact info:

cathymar@microsoft.com

http://www.csdl.tamu.edu/~marshall
http://research.microsoft.com/~cathymar