

If I only had 10 minutes to talk  
about personal archives:

**3 things** I've learned **+ 1**

---

Cathy Marshall

Microsoft Research, Silicon Valley

Personal Archives

Internet Archive, 16 February 2010

Thing 1: for individuals  
(families/friends/collectives),

**benign neglect**

is the most likely collection policy and  
stewardship technique...

stewardship is paved with...

# Bookmarks

## Saving files with a CD-RW drive



David Einstein  
Computing Q&A

My Windows XP computer came with a CD-RW drive but not a floppy drive. Is there a way to save files on a CD that's as easy as saving to a floppy?  
A: Saving to a CD is easy, and software is available to help you. You would not require a steep learning curve for a relatively straightforward set of things to do.

Microsoft Excel, which I use for my project management software, especially for small groups, are less than \$75. I recently switched Internet service providers from America Online to Comcast. With AOL, I could spell-check my e-mails but when I switched and began using Outlook Express, the spell-checker was no longer available. Is there any way to activate it?

Here's the deal: AOL has its own built-in spell checker, but you don't have to use it and it appears you don't, then you won't be able to use the spell-checker. All is not lost, however. One solution is to download a free spell-checking program for Outlook called Spell Checker for Outlook. You can find it at [www.mcafee.com](http://www.mcafee.com). Go there and download the spell checker.

Q: I want to have a PowerPoint presentation that shows slide after slide automatically, without the need to click to each new slide. Is there a way to do that?  
A: There is. With your presentation open in PowerPoint, go to the Slide Show menu and choose Section Transition. In the Advance section, click the box labeled "Automatically after," and choose the number of seconds you want slides to be on the screen. Then click Apply to All.

**TIP OF THE WEB**  
In a recent column, I discussed how to disable the feature that automatically turns Internet mail addresses into hyperlinks in Microsoft Word. A reader suggested a way to move individual hyperlinks to the Edit menu after Word chooses them. But...

start (\$129.95 from [www.projectkickstart.com](http://www.projectkickstart.com)) and TurboProject (\$99.95 for the standard version, \$49.95 for the Express version, from [www.immssoft.com](http://www.immssoft.com)). And if you happen to be a teacher or have a student in the house, you could get the academic version of Microsoft Project — normally a \$600 program — for less than \$75.

Q: I recently switched Internet service providers from America Online to Comcast. With AOL, I could spell-check my e-mails but when I switched and began using Outlook Express, the spell-checker was no longer available. Is there any way to activate it?  
A: Here's the deal: AOL has its own built-in spell checker, but you don't have to use it and it appears you don't, then you won't be able to use the spell-checker. All is not lost, however. One solution is to download a free spell-checking program for Outlook called Spell Checker for Outlook. You can find it at [www.mcafee.com](http://www.mcafee.com). Go there and download the spell checker.

**TIP OF THE WEB**  
In a recent column, I discussed how to disable the feature that automatically turns Internet mail addresses into hyperlinks in Microsoft Word. A reader suggested a way to move individual hyperlinks to the Edit menu after Word chooses them. But...

...good intentions

“...neglect can sometimes be an artifact’s best friend.”

- *G. Thomas Tanselle*

*“Statement on the Significance of Primary Records”*



reel-to-reel  
tape used to  
archive the  
rare vinyl  
record...

Multiple  
copies of a  
rare vinyl  
record



but 30 years later...

- *no reel-to-reel deck & media is decaying (wrong strategy)*
- *LP is readily available (long tail) (wrong policy)*
- *wrong about projected personal value (wrong assessment)*



benign neglect

as collection policy

digital belongings accumulate rapidly  
without regard for **value** or provenance

there are > **4.3 billion** personal photos on Flickr



# 2,626,042,804



Quicktake #26

but that's just the tip of the iceberg...

Facebook has *AT LEAST* three times that many (>15B/60B)  
and Image Shack has *AT LEAST* four times that many (>20B)

# accumulation



“[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 answered,



“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.”

value



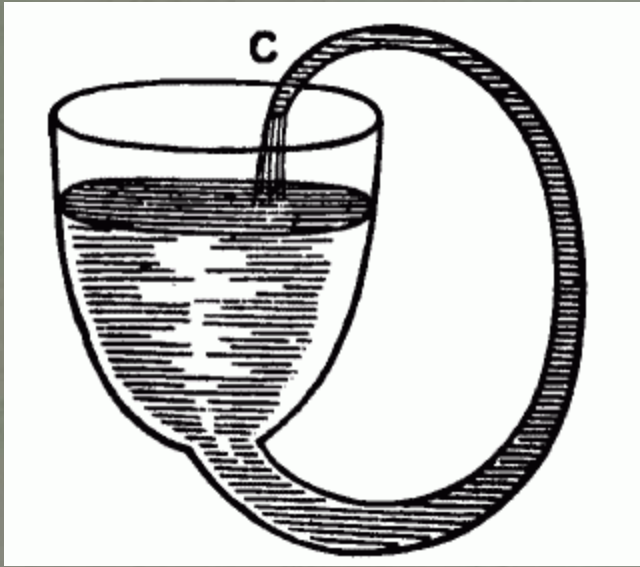
## *Loss as a means of culling collections*

“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.”

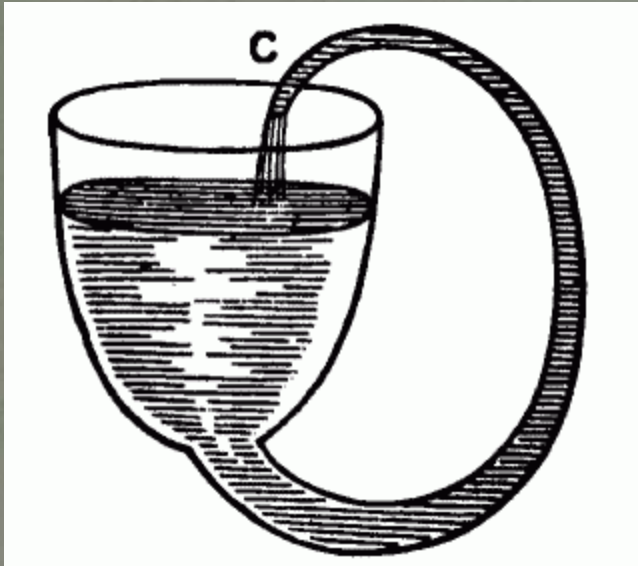
“If my hard drive was gone, it really wouldn’t bother me all that much, because it’s not something I need, need. I just thought it would be nice to keep it around.”

“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?”



It's easier to *keep*  
than to *cull*



It's easier to *keep*  
than to *cull*,

but it's easier to *lose*  
than *maintain*.

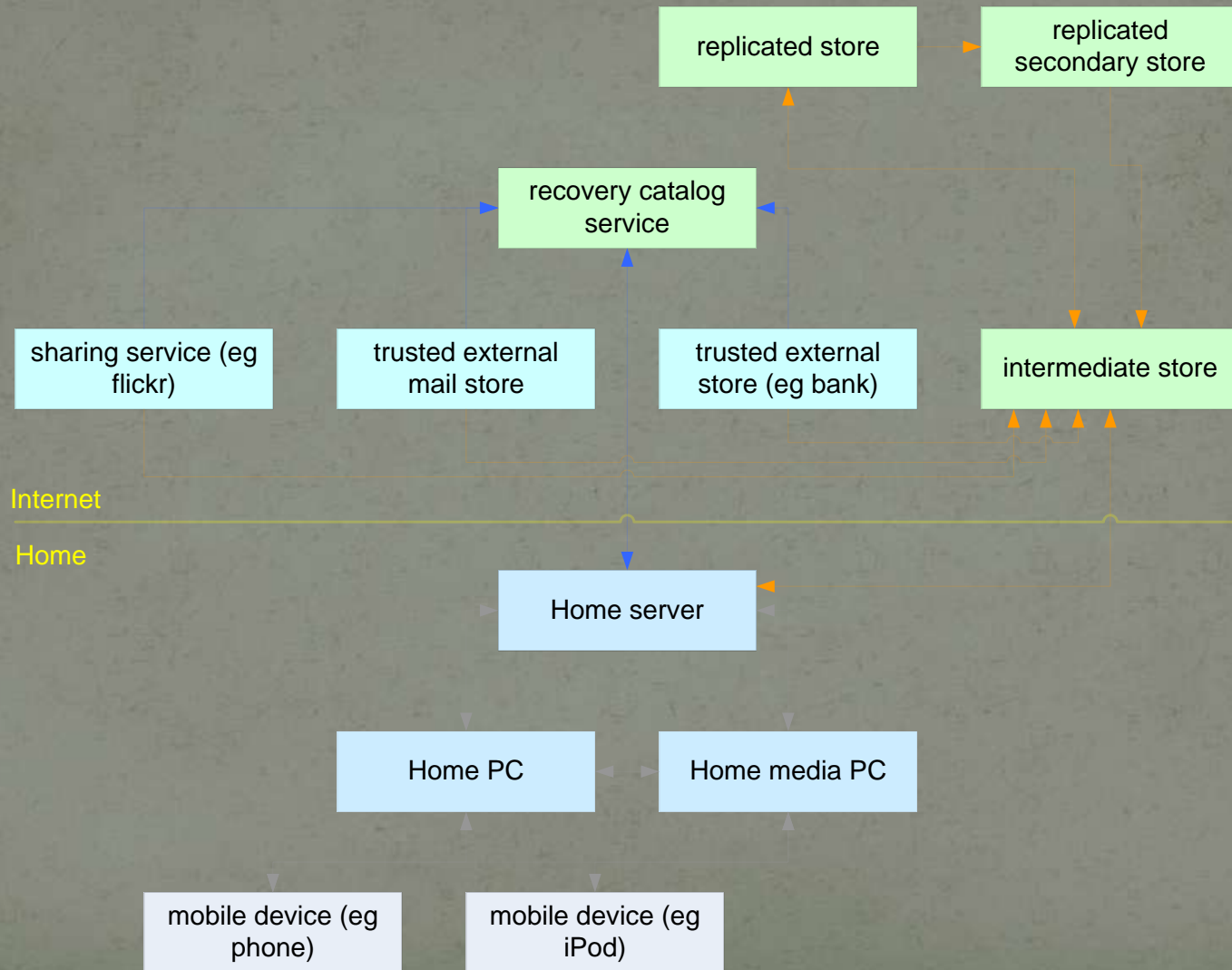
Thing 2: **no single** preservation  
technology/repository/file system/cloud  
store/next gen OS will "win"...

Today, there are two standard technical solutions: (1) shove everything into a big database in the cloud and decode it later (the **Oscar Madison** approach) ...



...or (2) safe storage and self-describing digital objects  
(the **Felix Unger** approach)

instead of either monolith, we'll be knitting together stores and services



# No single solution!

- catalogs, catalogs, catalogs
- think about your dental xrays
- now think about your high school yearbook
- new institutions/  
cultural expectations



Thing 3: Forget about digital originals  
or reference copies. Each copy takes  
on **a life of its own!**

[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.) zoomr (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



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

*data safety is a **side effect!***

This leads people to use circular reasoning about where the 'truth' lies...



We think of the local copy as archival (and it is in the sense that it's highest fidelity)

“The good thing about the photos is that there's always an intermediary step. I mean, like the photos go off of my camera onto my computer before they go up to Flickr. So I always have master copies on my PC. So that's why I don't care so much about Flickr evaporating.”

But... the web copies have been augmented with useful organization and metadata (e.g. tags and possibly comments)



“I didn't lose the pictures, but I was sorry that I had lost the collections and the organization, and you know. I'm sure I have the pictures somewhere still. But fishing them out and recreating it was not feasible.”

many **inexact** copies



400 x 300  
18k - jpg



400 x 300  
19k - jpg



500 x 374  
91k - jpg



450 x 337  
123k - jpg



550 x 412  
101k - jpg



413 x 309  
23k - jpg



400 x 300  
18k - jpg



500 x 360  
63k - jpg



550 x 412  
65k - jpg



400 x 300  
27k - jpg



550 x 412  
65k - jpg



450 x 324  
43k - jpg



500 x 360  
31k - jpg



500 x 344  
49k - jpg

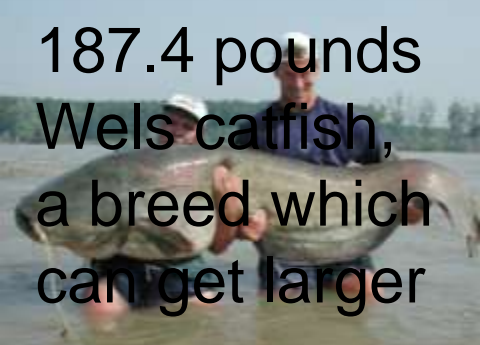


350 x 262  
16k - jpg



400 x 300  
19k - jpg

each with **a life of its own**

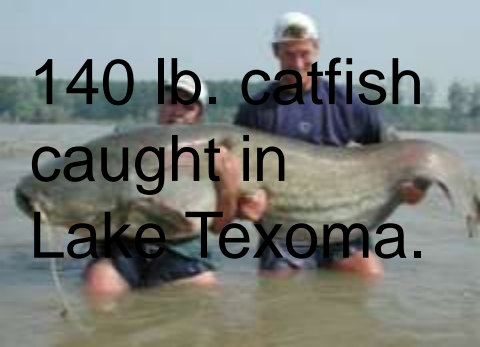


A Vietnamese catfish

187.4 pounds Wels catfish, a breed which can get larger

“that certainly is a big fish!”

Catfish from Mississippi, just shy of 646 lbs



in rphou's personal ftp directory

140 lb. catfish caught in Lake Texoma.

Photo of a Giant Catfish

from the Johnson Family Photo Album



A BIG Wels Catfish (187lbs) caught by Lucas Van Der Geest

it's not an American catfish

The Secrets of Catching Giant Catfish

World Record Catfish



Mekong catfish caught in the Mekong river (424 pounds)

“Two men catch catfish by sticking their hand elbow deep into the mouth of the fish”

“Fishermen with Giant Catfish: real image”

“It's called a Wel's Catfish. They get bigger than this!!!”

each with a *social* life of its own

# example: an animation stored on different social media sites. Same animation, different **social & descriptive metadata**



downloaded 387 times



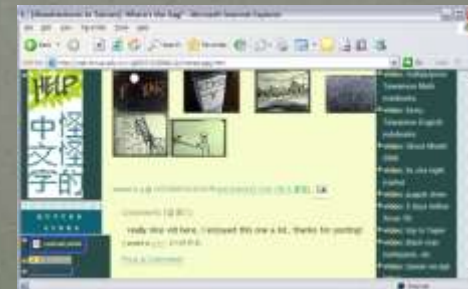
3,869 views, ★ ★ ★ ★ ★



45 views, no "likes"



viewed 245 times



"really nice vid here, i enjoyed this one a lot."

Thing 4: Given the first three things,  
there will be some interesting  
opportunities to take a fresh look at  
**searching and browsing**



# why it's **different**...

we might have forgotten it altogether: **re-encounter**

we've got some context and a rough idea of what we want: **faceted browsing**

we've got a good idea of what we want (even though it might not exist): **visualizations and desktop search**

*and then there's the **whoops** factor*



boiling the ocean?



NO! We're closer than we think...



contact info:

[cathymar@microsoft.com](mailto:cathymar@microsoft.com)

<http://www.csdl.tamu.edu/~marshall>

<http://research.microsoft.com/~cathymar>

blog—<http://ccmarshall.blogspot.com>

twitter—<http://twitter.com/ccmarshall>