or “people are people and things change.”

Cathy Marshall
Microsoft Research, Silicon Valley

Personal Digital Archiving 2011
24 February 2011
a familiar story...
in the fall of 2009, my laptop was showing signs of impending trouble. By December, things were dire.

Of course I did exactly what you’d expect.
I ignored it and soldiered on...
My laptop hasn't died yet. Nor have I backed it up. It's a game of chicken, and I think we know who's going to win.

6:47 PM Jun 28th via web

a week passes...

My laptop's still up on the hoist. "Needs a complete overhaul, ma'am" says the mechanic, dislodging Twix crumbs from the keyboard.

1,278,105,496,000.00 via web

another week passes...

My laptop's back from the shop. Everything's been replaced, except 1 DIMM and that new car smell.

1,279,152,732,000.00 via web

CATASTROPHE AVERTED!
this is the message that Windows showed me last night...
Set up backup
Your files are not being backed up.
Turn off messages about Windows Backup
So what’d I do? I turned off the annoying messages telling me I wasn’t backing up my computer...
at the same time as my laptop was failing, other things were going wrong too...
"After one look at this planet any visitor from outer space would say I WANT TO SEE THE MANAGER." -- William S. Burroughs

3:24 PM Jun 15th via web
1200 of my tweets had mysteriously disappeared
I want my darned tweets back. Please, twitter, pretty please. I promise to back them up next time.
3:26 PM Jun 15th via web

the next day...

Dear Twitter: Please give me my 1200 missing tweets back, or else I'll make vague empty threats in your general direction. Love, Cathy
11:50 AM Jun 16th via web

an argument breaks out...

@jerepick @gingdottwit @jonathankoren Some of my tweets are ephemeral; some (emphatically) are not! I'd give you examples, but they're gone.
12:01 PM Jun 16th via web in reply to jerepick

a happy ending!

12:03 PM Jun 16th via web
Over the last 6 months, I’ve documented no fewer than 3 ways to archive tweets.

ccmarshall  cathy marshall
Just when the DSM-V is ready to scrap Narcissistic Personality Disorder, along comes http://twournal.com/ (and I can’t wait to use it!)
13 Dec

THAT camp--a digital humanities meeting. Find out about it.
twitter archive "Twapper keeper"

On own server; can also export them to a csv file BackupMyTweets

McKemmish, “evidence of me” (1996)

1 billion camera phones in the world

According to @jubois, 35 hours of video uploaded to YouTube per week were so many embarrassing personal accidents.

What does it cost to endow a terabyte? Complex cost models like Cartwright—what about democratizing the cost. “Buy a brick” rather than a "brick house"?
not only have I not used them... because of changes in Twitter’s terms of use, these services may disappear.

---

**rdhyee** Raymond Yee
Twitter should be providing the archiving features that Twapperkeeper struggles to provide. [http://bit.ly/e31gbm](http://bit.ly/e31gbm)
7 hours ago

---

**ccmarshall** cathy marshall
Does this mean BackupMyTweets and Twournal don't comply with Twitter’s terms of service either? [http://bit.ly/e31gbm](http://bit.ly/e31gbm) (thanks @rdhyee)
22 Feb
The first part of the story—people do little to ensure the sustainability of the stuff on their computers—is old and familiar; the second part (the social media side) is new, but hardly surprising over the last 25 years, even Barbie has turned from aerobics to lifelogging...
Barbie’s just the tip of the iceberg...

- 2/3 of Americans store personal data in the cloud. 
  *Lee Rainie, Pew Internet, 7/2010*

- 48% of Americans are social networkers 
  *Lee Rainie, Pew Internet, 7/2010*

- ~70% of the world’s digital content is created by individuals 
  *UK Telegraph, IDC annual survey, 5/2010*

- In 2002 digital storage capacity overtook analog; as of 2007, information was 94% digital 
  *Martin Hilbert, USC Annenberg School of Communication*
so people are people and things change
The personal archiving territory has shifted dramatically over the last 6 years...

- consumer field study in 3 cities (2005)
  - surprises: islands of understanding amidst a sea of confusion + digital folkways + analogies to physical media don’t work = benign neglect and its side effects

- lost website recovery study (2007)
  - surprise: data loss often isn’t due to technology

- survey of social media ownership attitudes (now)
  - surprise: people have complex ideas about social media ownership and reuse
Things they said that surprised me...

- **consumer field study (2005)**
  
  “What? You’ve never heard of Whimit?” I. says with a note of disbelief in her voice

  “128 encryption yeah. We’d at least have that much [to store photos online]. 64 bits has been hacked easy.”

- **lost website recovery study (2007)**

  “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”

- **survey of social media ownership attitudes (now)**

  “[Reusing photos from the Web] is okay most of the time. The only time I would think it isn't would be when the main focus is of someone you don't know. Like when people email out the People of Walmart photos. Those are taken by people who don't know the person in the photo and posted.”
2005

benign neglect and its side effects
“[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.”

[do you ever get rid of digital stuff?]

“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.”
“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]”
the family archivist is not
the family photographer is not
the family IT support 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 registry is a battlefield...
personal digital archiving

==

archiving a personal digital collection
i.e. Michael Joyce archiving his own digital stuff

Gabby Redwine archiving Michael Joyce’s digital stuff
It’s easier to keep than to cull... but it’s easier to lose than maintain.
Is accumulation really a problem? Storage is cheap and getting cheaper. Why not just keep **EVERYTHING**?
In her 2010 dissertation defense, Christina Finneran posited out a connection between digital keeping and the TLC-popularized affliction of *hoarding*

*but does digital hoarding really exist? or is keeping everything just the normal state of affairs?*
i.e. is normal digital accumulation hoarding?
2007
our personal data is replicated online & has 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] ccm 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!
“nearly 70% of [reported data loss in the home] manifested itself in an inability to find information; by comparison ... 8 % [of loss was] due to hard drive failure.”

—British Library’s Digital Lives Synthesis
a recent social media drama...

“Apparently Wilhelm reported a Flickr user with an account that held ‘obviously stolen material’ to the company last weekend, but a staff member erroneously incinerated his [pro] account [containing 4000 manually ‘linked’ photos] instead of the culprit’s.”...

from a TechCrunch blog post earlier this month
copies take on lives of their own...

downloaded 387 times

3,869 views, ★★★★★

45 views, no "likes"

viewed 245 times

"really nice vid here, i enjoyed this one a lot."
People use circular reasoning

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. I’m sure I have the pictures somewhere still. But fishing them out and recreating it was not feasible.”
The saga of a giant catfish
A Vietnamese catfish

187.4 pounds Wels catfish, a breed which can get larger

“that certainly is a big fish!”

in rphou’s personal ftp directory

140 lb. catfish caught in Lake Texomma.

Photo of a Giant Catfish

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

140 lb. catfish caught in Lake Texomma.

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

A Vietnamese catfish

187.4 pounds Wels catfish, a breed which can get larger

“that certainly is a big fish!”

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

it’s not an American catfish

The Secrets of Catching Giant 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”

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”

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”

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”

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”
2010

social media (& social life) makes the **ownership** of all this stuff complicated!
ownership: a slippery slope

“Someone tagged a picture of me on facebook and i saved the picture and put it into my album because it was a nice picture.”

“my friend took a picture of my son and her son. i reposted so everyone can view from my family.”

“Earlier this week I downloaded a picture of a dog wearing a party hat for a story I was doing on my pet blog about an event coming up. It was a photo included in a press release by the store holding the event.”

“I couldn't find a photo I'd taken on a trip, which I wanted to use in a Facebook album, so I found a photo of the same landmark on someone's blog and republished it in my album.”
2. Sharing Your Content and Information

You own all of the content and information you post on Facebook, and you can control how it is shared through your privacy and application settings. In addition:
1. For content that is covered by intellectual property rights, like photos and videos ("IP content"), you specifically give us the following permission, subject to your privacy and application settings: you grant us a non-exclusive, transferable, sublicensable, royalty-free, worldwide license to use any IP content that you post on or in connection with Facebook ("IP License"). This IP License ends when you delete your IP content or your account unless your content has been shared with others, and they have not deleted it.
2. When you delete IP content, it is deleted in a manner similar to emptying the recycle bin on a computer. However, you understand that removed content may persist in backup copies for a reasonable period of time (but will not be available to others).
3. When you use an application, your content and information is shared with the application. We require applications to respect your privacy, and your agreement with that application will control how the application can use, store, and transfer that content and information. (To learn more about Platform, read our Privacy Policy and About Platform page.)
4. When you publish content or information using the "everyone" setting, it means that you are allowing everyone, including people off of Facebook, to access and use that information, and to associate it with you (i.e., your name and profile picture).
5. We always appreciate your feedback or other suggestions about Facebook, but you understand that we may use them without any obligation to compensate you for them (just as you have no obligation to offer them).

You retain your rights to any Content you submit, post or display on or through the Services. By submitting, posting or displaying Content on or through the Services, you grant us a worldwide, non-exclusive, royalty-free license (with the right to sublicense) to use, copy, reproduce, process, adapt, modify, publish, transmit, display and distribute such Content in any and all media or distribution methods (now known or later developed).

Tip This license is you authorizing us to make your Tweets available to the rest of the world and to let others do the same. But what's yours is yours — you own your content.

You agree that this license includes the right for Twitter to make such Content available to other companies, organizations or individuals who partner with Twitter for the syndication, broadcast, distribution or publication of such Content on other media and services, subject to our terms and conditions for such Content use.

Tip Twitter has an evolving set of rules for how API developers can interact with your content. These rules exist to enable an open ecosystem with your rights in mind.

Such additional uses by Twitter, or other companies, organizations or individuals who partner with Twitter, may be made with no compensation paid to you with respect to the Content that you submit, post, transmit or otherwise make available through the Services.
Facebook: “You grant us a non-exclusive, transferable, sublicenseable, royalty-free, worldwide license to use any IP content that you post on or in connection with Facebook... This IP License ends when you delete your IP content or your account unless your content has been shared with others, and they have not deleted it.”

Twitter: “You retain your rights to any Content you submit, post or display on or through the Services. By submitting, posting or displaying Content... you grant us a worldwide, non-exclusive, royalty-free license (with the right to sublicense) to use, copy, reproduce, process, adapt, modify, publish, transmit, display, and distribute such Content in any and all media or distribution methods (now known or later developed).”

Flickr: “With respect to photos... you submit or make available for inclusion on publicly accessible areas of the Yahoo! Services... the license to use, distribute, reproduce, modify, adapt, publicly perform and publicly display such Content on the Yahoo! Services soles for the purpose for which such Content was submitted or made available. This license exists only for as long as you elect to continue to include such Content on the Yahoo! Services and will terminate at the time you remove or Yahoo! Removes such Content from the Yahoo! Services.”
overarching research questions

- **limits**: what do people feel they can and can’t do with media they didn’t create themselves?

- **reciprocity**: can you do the same thing with my stuff that I do with yours?

- **reuse**: What constitutes fair use of stuff you find on the Web?

- **scale**: do people feel that public institutions should be able to archive and provide access to publically accessible social media?
reciprocity: reuse is rife with contradictions...

- funny v. making fun of
  “if someone is in the background of the photo and is tripping over a friend's leg or something on the ground and fall flat on their face, I don't think it's ethical or right to be able to post it.” [R182]
  “i sent out an email of ugly prom pics” [R235]

- altered v. fraudulent

- children v. minors
  “I think [reuse] endangers kids..., makes kidnapping and abuse more accessible, makes life of perverts better. I don't like that anyone can look at photos of my nieces on Flickr or Fbook. I think kids' images should be especially regulated” [R082]
  “I downloaded a pic of a cute baby i saw online to use as my desktop background.” [R011]
reuse covers complex ethical and moral terrain

- law
  
  "the web is public domain. if you post a picture (with or without a copyright watermark), it's going to be seen and recycled by others. if you don't want your picture or face to possibly end up being seen by future bosses, kids, friends, parents whatever, then don't post it." [R153]

- technology
  
  "I only post photos that feature the ‘share’ button for Facebook built in to the blog." [R140]

- creation context
  
  "Personal photos from someone's personal album should probably not be reused. Perhaps if it is picture someone takes of a flower, or a sunset, and you find it pretty and want to save it, then that's fine..." [R227]

- reuse context
  
  "...If the user will be profiting from your photo or using it for personal use. If it's for personal use then it would be okay. If it's on a website that makes money but it's not money coming from your photo but your photo makes the website look better then I think that would be okay also." [R207]

Reuse in practice may be mediated by need, the nature of the material, history (e.g. how much it has already ‘been around’)
The Library of Congress is acquiring the public archive

<table>
<thead>
<tr>
<th>Statement</th>
<th>Twitter (mean)</th>
<th>Flickr (mean)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Library of Congress can give everyone access to the archive.</td>
<td>4.22</td>
<td>4.14</td>
</tr>
<tr>
<td>The Library of Congress can give researchers access to the archive.</td>
<td>4.72</td>
<td>4.65</td>
</tr>
<tr>
<td>The Library of Congress can give everyone access to the archive after 50 years has passed.</td>
<td>4.59</td>
<td>4.93</td>
</tr>
</tbody>
</table>

- Most positive response for Twitter? Access by researchers.
- Most positive response for Flickr? Access in 50 years.
What should we make of this series of changes?
“No, I’m not backing up our files—I’m just assuming that the F.B.I. is making copies.”
The REST ASSURED Disposal Service
(a Man Show conceit)
Many years ago, I appeared naked on the cover of a magazine. Well, I wasn't completely naked. I was covering my crotch with a magazine...

My then-girlfriend was starting a literary magazine that would also be accessible online and she had a specific idea for the cover of the inaugural issue -- me, naked, looking terrified, with a magazine over my crotch, standing in a New York City subway car. It would be funny, interesting and eye-catching...

[My relationship with my grandmother, who was shocked by the photo] certainly outlasted the magazine, which is since defunct, and even the photo, which is now a casualty of **online neglect**, nowhere to be found...

---Michael Newman, Open Salon blog
kid 1: “They should just save Facebook. That is our generation’s scrapbook, yearbook, Guinness World Record…”

kid 2: “But obviously you can’t save everything”

LoC narrator: “But the truth is, digital information will survive only as long as someone takes care of it.”
credits

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

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

- Shared information ownership study collaborator: Frank Shipman (TAMU)
contact info:

cathymar@microsoft.com

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

blog—http://ccmarshall.blogspot.com
twitter—http://twitter.com/ccmarshall