
Benign Neglect and the Institutional Archiving of Personal Ephemera

Catherine C. Marshall

Texas A&M University
College Station, TX 77843, USA
ccmarshall@cse.tamu.edu

Paste the appropriate copyright/license statement here. ACM now supports three different publication options:

- ACM copyright: ACM holds the copyright on the work. This is the historical approach.
- License: The author(s) retain copyright, but ACM receives an exclusive publication license.
- Open Access: The author(s) wish to pay for the work to be open access. The additional fee must be paid to ACM.

This text field is large enough to hold the appropriate release statement assuming it is single-spaced in Verdana 7 point font. Please do not change the size of this text box.

Each submission will be assigned a unique DOI string to be included here.

Abstract

This position paper argues that some portions of Facebook, collected at intervals, can be usefully archived in a manner on par with census data. I first explain why benign neglect is a likely model for personal digital archiving, leaving peoples' digital legacies to chance. I then explore users' attitudes to archiving Facebook content (using an automated mechanism), and their objections to the corresponding institutional effort to collect and preserve personal ephemera. Finally, I raise issues that will need to be addressed before this effort can be undertaken.

Author Keywords

Personal digital archiving; PIM; Facebook; social networks; social media; benign neglect.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous

Introduction

The long-term viability of our personal information depends partly on our commitment to its stewardship, but also in a large measure on chance, on the outcome of a lifetime of benign neglect. Increasingly, each of us is responsible for creating, recording, gathering, curating, and most of all, *accumulating* large

heterogeneous collections of digital materials. Some of the material is thought of as highly personal (e.g. email) and is kept as a private resource. Other content is maintained socially (e.g. feeds circulated within the confines of social media platforms) or is published openly on the Web (e.g. blogs, YouTube videos). Still other content—records, mostly—are held privately and created and accessed by the institutions that generate the records (e.g. medical or financial records).

We can slice and dice this type of personal information into many different categories, but the important thing to realize is that it is difficult for most people to manage this much structured and unstructured content that is distributed among many services, devices, and forms of storage media, especially as time passes.

It is unusual for people—including numerous study participants [3,2] —to sustain normal curatorial effort over extended periods of time. This is not to say that people never curate personal information. They may devote significant time and resources to collecting and curating their digital belongings during discrete periods of their lives (say, after purchasing a new device, or when a major life event has occurred, a wedding, say, or the birth of a child). Seemingly negative events (e.g. a break-up or divorce) may trigger substantial deletion and culling [6].

Yet, if there's one thing that fieldwork has shown us, it's that in spite of good intentions and aspirations to "be more organized" or to "stop being such a pack rat", participants might not know how to systematically preserve digital belongings; they may not have time to curate them; and they may be unaware of (or unwilling to invest in) services to help them address this

problem. Furthermore, automatic services may merely compound the problem: people often forget how to login to a service, or they may forget that they've created an account on a service [1]. ***Often people lose digital materials through the same actions they use to ensure their safekeeping.***

Benign neglect has several important implications:

- Digital belongings will continue to accumulate through most peoples' lifetimes, without regard to their value, often until they are lost or forgotten.
- What survives individuals' lives is left largely to chance. Because bits are less visible than their physical counterparts, there is little sense of what people have and where they've kept it.
- People are generally just as concerned with inadvertently leaving something damaging behind (e.g. porn, secrets) than they are with preserving a personal legacy. For example, a vegetarian may be disconcerted by the thought of a photo of herself as a young adult in which she was eating a hamburger.
- Generations don't have reciprocal perceptions of an item's value. That is, people leave things to their children that their children don't want. Similarly, they discard items that the children may value.

Ephemera and personal digital belongings may have enduring value as historical and cultural artifacts. Although much is made of quantitative data collected over the years (say, in service of historical demography), there are also good reasons to preserve ephemera and non-official records too. Historians, archivists, and social science researchers are well aware of the value of this type of seeming ephemera.

Four implications of benign neglect

Uncontrolled accumulation without regard to value

What is kept is a matter of chance

People are more concerned with deleting than keeping

Generations don't have reciprocal perceptions of value

Four primary user objections to institutional archiving of Facebook content

Right to identity control

Potential for malicious use of personal data

Expectation of veracity of materials held by public institutions

Low societal benefit relative to costs

Even amateur genealogists have experienced the joy of discovering a missing bit of data in a century-old city directory.

In this position statement, I will argue for the need to collect personal ephemera—particularly content stored and shared in social media platforms like Facebook—in embargoed, access-restricted institutional archives (possibly in dark archives to put the material out of reach of overly ambitious governments). Such an effort should begin soon, before the inevitable loss occurs.

Attitudes toward Personal Archiving

Users of social media platforms like Facebook are notably nonplussed by proposed personal archiving efforts. Even if Facebook archiving were performed on their behalf (without requiring extra effort on their part), users seem largely uninterested.

Even if users are gently reminded of all the time and effort they've put into sharing stuff on Facebook (eg. Zhao and Lindley have performed studies that rely on exercises to make users more aware of the value of their Facebook content [7]), they usually are skeptical that they harbor an unrealized desire to keep it forever. First, they perceive the existing service as a long-term store as it stands: everything they have ever put there is available on an on-going basis and there seems to be little reason to be concerned that it will go away anytime soon. Second, most discrete items—for example, photos and videos—are stored other places too (for example, on users' phones or in cloud backups). Finally, and most importantly, users question the persistent worth of much social media content; its value was in the moment, as part of communication and self-presentation.

Even if a hypothetical integrated personal archive were constructed effortlessly—content was harvested from its many online habitats (not just Facebook) and stored safely for easy future access—people sigh at the overall cluelessness of this idea. The incipient loss of context inherent to such a personal archive makes it seem even more useless. [5]

Attitudes toward Institutional Archiving

It is inarguable that eventually even the most popular commercial services will fall victim to changing fortunes. For example, MySpace, once an ascendant social networking site, has become a niche service. Other services like Geocities are gone for good. It is our good fortune that amateur archivists made copies of the final state of the popular site.

Yet institutional archiving efforts have met with little applause. Many of us remember the flap associated with Twitter's donation of its public feed to the Library of Congress: it is easy to believe that even more people will question the virtue of a Facebook archive. When we've posited the need to archive Facebook [4], sometimes even close colleagues turn hostile. "You're going to ruin Facebook for me," they say. "I'll never use it again." Others don't object only because they sense they are powerless in the face of complex license agreements they have signed. At best, some people sigh and shrug their shoulders as if to say "what're you going to do?" because they believe the content they post to Facebook is so anodyne that it is unimportant what happens to it.

Frank Shipman and I have performed a series of studies of social norms for ownership and control of digital content. From our results, it seems like the main

Issues to be resolved

Facebook's role in content donation

Basic content to form the backbone of such a collection

Utility of a main profile-and-network collection strategy

Curation policies

Collection intervals

Access policies (embargoes, limits on access, aggregation, and so forth)

Data protection strategies

Donor agreements for persons desiring a broader legacy

Resources necessary to undertake this effort

objections to institutional archiving of the PIM content held by Facebook can be characterized in four ways:

- *Institutional archiving interferes with individuals' rights to control their online identity.* This perceived right is often mischaracterized under the rubric of privacy. Instead, users are worried that their self-presentation won't reflect the latest version of themselves.
- *Malicious uses of this sort of personal data outweigh the good ones.* There's a fear (and it doesn't seem to be unwarranted) that the data will be used maliciously by individuals, commercial interests, or governments. This worry seems like it might be mitigated by limiting or deferring access.
- *Content veracity and timeliness will not be guaranteed.* This may be an expectation set by the availability of large-scale resources like Wikipedia. Facebook may seem inaccurate in comparison.
- *There is no particular societal benefit to using public resources this way.*

The first and second object are best countered by a well-considered access policy. Embargo periods, data aggregation, and limited access to qualified persons all mitigate risk somewhat. Storing the content in a strong dark archive comes closer to eliminating risk, although this strategy is probably overkill. The third objection and fourth objections, veracity and low societal benefit, would need to be countered by a public education campaign.

Issues

This is a controversial proposal. Yet historical research will benefit from such a collection. What's in the

collection requires substantial thought to balance utility, privacy, expectations, and policy. For example, many people regard their profiles as something they've published. Hence account profiles and the Facebook social network may be the basis for this collection. Social networks are not static; appropriate collection intervals should be determined. Access policies need to be worked out to mitigate risks. Encryption policies must also balance risk of data loss with risk of early dissemination. Donor agreements may be developed to enable people to contribute content beyond the basics.

References

1. John, J. L., Rowlands, I., Williams, P. and Dean, K. *Digital lives: personal digital archives for the 21st century*. V. 0.2 (3 March 2010).
2. Lindley, S., Marshall, C.C., Banks, R., Sellen, A. and Regan, T. Rethinking the web as a personal archive. In *Proc. WWW 2013*, pp. 749-760.
3. Marshall, C.C., Bly, S., and Brun-Cottan, F. The Long Term Fate of Our Personal Digital Belongings. *Proc. Archiving '06*. Soc. Imaging Science and Technology, Springfield, VA, (2006), pp. 25-30.
4. Marshall, C.C. and Shipman, F.M. An Argument for Archiving Facebook as a Heterogeneous Personal Store. *Proc DL'14*.
5. Marshall, C.C. and Shipman, F.M. Exploring the ownership and Persistent Value of Facebook Content. In *Proc. CSCW'15*.
6. Sas, C. and Whittaker, S. Design for Forgetting: Disposing of Digital Possessions After a Breakup. *Proc CHI'13*.
7. Zhao, X. and Lindley, S. Curation through Use: Understanding the Personal Value of Social Media. *CHI'14*, pp. 2431-2440.