

# An Argument for Archiving Facebook as a Heterogeneous Personal Store

Catherine C. Marshall  
Microsoft Research  
1065 La Avenida  
Mountain View, CA 94043  
1-650-693-1308  
cathymar@microsoft.com

Frank M. Shipman  
Department of Computer Science  
Texas A&M University  
College Station, TX 77843-3112  
1-979-862-3216  
shipman@cs.tamu.edu

## ABSTRACT

A decade ago, the locus of activity for our digital belongings—photos, email, videos, documents, and the like—was on our personal computers. Now the situation is different. Not only is personal media born-digital, it may also spend its entire life stored online in social media services and cloud stores, and locally on portable devices. Studies have revealed that most people lack the requisite skills to archive their digital belongings, regardless of where they are stored; furthermore people value the context offered by these large-scale, socially intertwined online stores. So why not archive the contents of a major social media service like Facebook to ensure the permanence of a meaningful portion of peoples' personal digital belongings? Rather than being delighted by this idea, participants in a study of digital ownership have expressed squeamishness about institutional efforts to archive social media: Facebook is not only viewed as private and vulnerable to violations of content ownership, but also as lacking long-term value. However, measures such as data embargoes, aggregation, and permissions mitigate participants' fears and objections to some extent. In this paper, we will use an example of biographical research, coupled with the results of a recent study, to argue that Facebook should be archived by a public institution.

## Categories and Subject Descriptors

H.3.7 [Information Storage and Retrieval]: Digital libraries – Collection, Dissemination, User issues.

## General Terms

Design, Reliability, Human Factors, Legal Aspects.

## Keywords

Social media, Facebook, social networks, archive, historical research, personal information.

## 1. INTRODUCTION

As of January, 2014, Facebook had become a locus of social interaction and media sharing for 1.31 billion users in different parts of the world [29]. Curating one's Facebook profile and managing one's Facebook content has emerged as an important aspect of profection, the projection of one's identity into the

virtual world [13]. In this paper, we focus on the creation of an institutional archive of Facebook content, including the conditions under which it may be created, and the provisions under which it may be accessed and used, both now and in the future.

Early studies suggested that people are not necessarily capable of or motivated to maintain an on-going personal archive of their digital belongings, regardless of whether the items are stored locally or in the cloud [14]. Although services [18] and outreach programs [28] have been developed to help people create personal archives, evidence suggests that one important way that people lose their digital assets is by forgetting about them: they either fail to access the material, forget how to access the material, or forget about the existence of an account entirely [9,14].

In a recent study, Lindley and her colleagues found that participants rely on the specific online service where digital belongings are stored to give the assets meaning and context [12]. An integrated personal archive that aggregates items stored in different services would not be the same as preserving the items in their original context: a photo stored in a Facebook album or used in a Facebook profile has a different meaning to its owner than the same photo stored locally or in Flickr or Picasa. Furthermore, the study revealed that some social media types and services are seen as more archival than others; in particular, participants think of material shared on Facebook as transient, as part of a communicative action, not as something they would find meaningful in a personal archive. Finally, the study shows that time takes its toll on social media: there is a distinct tension between *keeping digital media* and *keeping digital media up to date*. For example, to signal abandonment of an account, its contents may be partially destroyed.

All of these factors seem to argue against archiving less curated, more transient and fluid social media services such as Facebook. Yet we believe there to be utility to such an archive, especially since many other comparable resources—e.g. city directories and paper correspondence—are disappearing in the Internet era. In this paper, we discuss the results of a study that explores which aspects of Facebook have value to its users, and what participants' concerns are when an institutional archiving effort is posited. We then describe a biographical research effort that relies on a number of different types of archival data stores, and use it in conjunction with our study results to suggest an approach for archiving a controversial resource like Facebook.

## 2. STUDY

The study we describe in this section is part of a family of related studies. Most particularly, we are pursuing the line of reasoning introduced in [16], which describes participants' attitudes to

hypothetical institutional archiving efforts based on different public media collections, including tweets, Flickr photos, Amazon book reviews, YouTube videos, educational recordings, and podcasts. A subsequent study reveals the relative sensitivity of data collected during multiplayer games [24]. Each of these studies elicits participant attitudes using a similar scenario: the Library of Congress (as a proxy for a large public institution) acquires a public collection of the media type in question. We tested three access conditions for each media type: immediate public access, delayed public access (after an embargo period of 50 years), and immediate researcher access.

Study results show that in the case of immediate public access, participants are likely to be concerned about ownership and intellectual property entailments: the four media types that are informally shared rather than formally published (tweets, photos, reviews, and game-related data) elicit substantially more negative reactions than the three published media types (videos, podcasts, and educational recordings). Although limiting immediate access to researchers reduces negative reactions, it does not appear eliminate privacy concerns: on the contrary, more personal media types (tweets, photos, and game-related data) still elicit a negative response. A 50-year embargo further reduces and evens the concerns out so they are similar across the board. The across-the-board negative reactions from about 20% of participants is likely to reflect fears that the content will be out of step with the contributors' current online identity.

This time, we turned our attention to Facebook. We recruited 250 participants on Amazon Mechanical Turk who have been regular Facebook users and who have established at least a 95% acceptance rate for their prior work. We used the methods and conservative screening rules described in [17]; data from 6 participants was discarded because it violated our reliability standards. In general, this method of recruiting has yielded good participation from a segment of regular Internet users who have integrated digital media and communication into their everyday lives.

**Table 1. Participants' common Internet activities**

Internet activities	# (% of 244)
email	241 (98.8%)
social networking	239 (98.0%)
shopping	217 (88.9%)
video sharing	168 (68.9%)
IM/chat	110 (45.1%)
Twitter	108 (44.3%)
videoconferencing	99 (40.6%)
photo sharing	87 (35.7%)
MMOGs	73 (29.9%)

Evidence from the demographic portion of the study suggests that the participants represent a broad spectrum of Facebook users. Of 244 participants, 52% are male. About a quarter of the participants are students. Only 8% have not attended college. Most (68%) have more than 10 years of experience on the Internet. Although the largest portion of the group were born in the 1980s (48%), almost a quarter are younger, and more than 10% were born in the 1960s or earlier. We also ask in multiple ways what else the participants do on the Internet to get a better picture of their online involvement and the role Facebook plays in their lives; Table 1 summarizes these common activities. An open-ended follow-up question shows that social networking is an important component

of participants' online activities, but it has not necessarily replaced other forms of communication (e.g. almost 99% still use email and 45% use IM or chat) or other venues for media sharing (69% use video sharing forums like YouTube and 36% use photo sharing services like Flickr or Instagram).

In this paper, rather than presenting results from the study's full range of scenarios, we focus on the scenarios and open-ended questions salient to our argument about the institutional archiving of Facebook.

### 3. RESULTS

#### 3.1 Use of Social Networking Platforms

Although there are now many social networking services—places to share media and connect with friends—in this study, we are focusing on general-purpose social networking platforms rather than specific media-sharing services. To find out which services participants relied on, we asked them two related questions: which social networking sites they use and whether they use any of these sites in an unusual way. Open-ended responses about unusual uses were coded and analyzed separately. All but 2 of the participants said they currently used Facebook; those two explained that they had recently deleted themselves from the service. Other services were secondary: 83/244 participants also used LinkedIn, and 73/244 also used Google+. A few listed media-type specific sites like Instagram (8) and Tumblr (6), or discussion forums like Reddit (5), but it was clear that Facebook is the dominant social networking service for the participants. Although the majority of participants (145/244) use more than one social networking service, a substantial number of those responded that they primarily used Facebook for their everyday social networking: 98 said they only use Facebook.

Table 2 shows the number of active social networking accounts that participants specified (the accounts may be on a single service). This practice shows two things: (1) Each participant is likely to have at least one profile that represents the 'real' them and (2) It may be necessary to go beyond the public portion of Facebook to see the real component of their digital footprint when participants have "friends-only" private personas [25].

**Table 2. Active social networking accounts per participant**

# active accounts	# participants	%
1	93	38%
2	82	34%
3	51	21%
4	7	3%
5	8	3%
6	3	1%
<b>Total</b>	<b>244</b>	<b>100%</b>

Most of the unusual uses that participants reported fell into four categories, all of which meant that the participant had to maintain multiple profiles. These results reinforce the existing findings about establishing multiple online identities [4]. Thirteen participants said they established multiple profiles to separate business and personal use. Several participants used this strategy to control advertising and spam; e.g. SN043 said...*I have two Facebook profiles (one real, one for entering contests and signing up for spammy stuff)*. Nine others found themselves wanting to post different material to different audiences. But maintaining separate identities is not seen as a panacea. SN194 said, *I have multiple profiles... But since both profiles can be found by anyone searching, I don't put anything inflammatory on that one either.*

Five participants explicitly used a divided social media identity to either seek privacy and/or maintain anonymity (sometimes for nefarious purposes). For example, SN165 admitted, *Like many users, I use multiple, fake profiles on Facebook for trolling purposes (messing with friends, anonymously commenting on sites, or being annoying in general when bored)*. This confirms our supposition that the participant's real profile may be off-limits behind privacy settings and that most people play it straight with at least one of their Facebook accounts. This aspect of Facebook data will become an important part of our argument later on.

Beyond digital identity, what is the function of Facebook and Facebook-like accounts to their owners? Knowing this will help us figure out what the participants imagine ownership of Facebook content to entail. Thus we asked participants what they'd done recently, given a menu of possibilities (see Table 3).

The most common Facebook activity is commenting: 82% of participants had written a comment during the previous week, squarely reflecting Facebook's communication and keeping in touch function. The second most common activity was photo sharing (55%); this involves curating personal content. Next was a purely curatorial activity, sharing found content (an article, joke, or video). Fewer than half had sent mail, and fewer than a fifth had edited their profiles. Responses tabulated in Table 3 demonstrate that participants had divided their efforts between communicating and curating material during the previous week.

**Table 3. Facebook use during the week prior to the survey.**

How Used	# participants	%
Commented	199	82%
Shared photos	133	55%
Posted article, joke, or video	128	52%
Sent mail	111	45%
Edited my profile	45	18%
Logged in, but did something else	21	9%
Haven't logged in/blank	5	2%

As a secondary ownership probe, we asked participants what they would expect to take with them if they moved to another social network. This response helps us identify content participants think of as not strictly transient. If the service is primarily used as a communication tool, we might expect the data to be viewed as mostly ephemeral, with the exception of a person's contacts. On the other hand, if Facebook serves as a repository, we might expect people to focus on their photos and messages. In that case, the content's value is what makes the service sticky.

With these constraints in mind, Table 4 shows that over half of the participants (54%) wanted to take their photos with them; some had considerable investment in them. SN217 said, *I have over 500 videos and thousands of photos on Facebook if I were to move I would expect to take those with me*. They also felt ownership of photos they'd taken or were featured in, in spite of the EULA most realized they had signed: *[I'd expect to take]... tagged posts and photos containing my name*. [SN191] Some anticipated a fight over ownership: *Any content I create and I purposely only upload photos with watermarks on them so even if facebook steals a photo everyone knows it*. [SN061]

The other Facebook element to consider is the social network, the connections one maintains. 85 participants (35%) mentioned wanting to take their contacts (or friends) with them. That said, a substantial proportion of those who thought of Facebook as a communication tool explained that they should be able to readily

rebuild what they had on Facebook, since they should know who their friends were: *...A good bit of my contacts on Facebook I am no longer in contact with are not necessary*. [SN182]

Given other explorations of sources of value in social networks [26], it is not altogether surprising that participants were less interested in taking content such as their wall posts and messages with them. In fact, profile information seems to be more valuable, if simply because it's so time-consuming to rebuild it (e.g. profile-based 'likes' may represent a significant curatorial effort); still, only 39/244 (16%) singled out their profile as something they wanted to move, and 9 others were interested in keeping their profile photos. If people are supplanting conventional email with Facebook, they must be saying little that they want to keep. Only 9 participants wanted to move their messages.

Finally, we can look at the extremes, participants who expected to take everything with them and participants who expected to leave it all behind. While only 24 participants (10%) expected to take everything, many of these responses were the most passionate: *Everything, All of my photos and photos that I have been tagged in. All of my mail and messages. All of the comments on my pictures and things that I have shared. All of my links that I have shared on my wall since joining. Anything else on my wall others have shared, pictures, links etc... And all of my game activity and causes information*. [SN188] A few even thought that the weight of their existing content was enough to keep them from moving, explicitly referring to the content's archival value.

At the other extreme, 50 participants (20%) looked at a move as a fresh start, or who thought it wouldn't be that hard to rebuild their social network (and that Facebook owned what they'd stored there anyway). SN203 returned to the communication tool theme, saying that he would expect to take *Nothing if I can't remember [sic] your name your [sic] not that important to keep*.

**Table 4. What participants expected to take with them from Facebook. Participants were allowed to click multiple options.**

Content type	# of responses	%
Photos	131	54%
Contacts/friends	85	35%
Nothing	50	20%
Profile information	39	16%
Everything	24	10%
Statuses	14	6%
profile photos	9	4%
Messages	9	4%

Finally, to anticipate destructive changes in the network (items that would appear and disappear) and objections to the effort, we asked participants an open-ended question about what they'd last removed from Facebook and why. Table 5 shows the social media elements participants reported as the last thing they had removed. Photos are the most common content type participants say they've deleted; 89 participants (36%) report they have removed photos recently, followed by posts or comments (22%).

Taken together, there are four primary reasons participants have removed Facebook content, with the first two being the most common: (1) to curate the material in one's accounts and keep the account itself tidy (*I hard-deleted my old Facebook profile in August and created a new one in its place. I had too many photos and too many friends, and I wanted a 'fresh' profile/account*. [SN193]); (2) to maintain one's online reputation (*I removed several alcohol related pictures when I started to look for*

teaching jobs. [SN024]); (3) to respond to an emerging understanding of privacy ([I] removed [my] birthday as it is often a question on a multifaceted security question and answer [SN011]); and (4) in reaction to a revelation about the service itself (After it came out that my content was being monetized with no reimbursement to me, I deleted my entire account. [SN061]).

**Table 5. Recent episodes of content removal by type.**

Content type removed	#	%
Photos	89	36%
N/A or nothing	53	22%
Posts/comments	53	22%
Profile elements	21	9%
Unspecified or 1-off	20	8%
Entire account	11	5%
Friends	8	3%
Links	6	2%
Game- or app-related auto posts	6	2%

Participants were fairly sophisticated about ways to remove or hide unwanted content: not only did they delete items they'd written themselves, like comments; they also fiddled with privacy settings, untagged posts and photos, and sometimes removed and replaced entire accounts so they could start over. Participants also managed their "likes"; these personal elements can appear unexpectedly in search results [15].

We also asked participants what they *wish* they could've removed from a social network; this may reflect the character of items that they would ask to be withdrawn from an institutional Facebook archive. Responses indicated that participants mostly feel like they are in control of their social network content; 56% (137/244) say they have no need to remove anything beyond what they are capable of removing. If we compare this to the answer to previous question, we can see that it's not so much that people always post the right thing (191/244, or 78%, say they have removed something), it's simply that they feel capable of removing content they don't want to keep in the service.

The most common exception is photos; almost one quarter (58/244, or 24%) say they would have been happier if they could have removed photos. Most of these were cases where the photo was simply unflattering. In most of reported incidents with posts (26/244, 11%), the damage has been done; e.g. one participant was fired for a public comment. Perceived privacy breaches (when someone else is responsible for compromising the participant's privacy) and legal issues also came to light. E.g., SN166 complained about others not maintaining her privacy

standards: *I don't like ... having people with lower security than me able to post photos of my kids without having to ask my permission* and SN088 reported *...An investigator...forced me to login on his own computer to look at the profile (although this is actually illegal in California and I should not have allowed it).*

### 3.2 Institutional Archiving of Facebook

We used two methods to explore participants' attitudes to the institutional archiving of Facebook: first we presented an institutional archiving scenario, followed by five hypothetical conditions for accessing that archive (described below); second, near the survey's end, we asked an open-ended question about whether it would be okay for a public institution to archive Facebook. The scenario and its hypotheticals primed participants to think about the possibility of this happening; the open-ended question gave them a chance to react to the archive's objectionable aspects and ways around them.

We developed the Facebook scenario mindful of the separation between public Facebook data (profiles and pages anyone can see) and data that is only visible through explicit relationships; we were aware that if we included private Facebook data, it was apt to produce a negative skew on the results. We also stipulated Facebook's importance as a cultural artifact so the question would be less of a referendum on Facebook's overall value (which we know is questioned). For the sake of comparison with the other 7 surveys, we adopted the same three basic conditions: access to the collection by the general public; a 50-year embargo on access; and access limited to researchers. We also added two conditions: aggregate access (the data would not reveal individual identities) and user-controlled archiving, which would allow users to specify what is kept. Figure 1 documents the scenario and its five hypotheticals.

Given the breadth of its adoption and the amount of content people have shared using Facebook, Facebook has become an important cultural artifact. Facebook's Board of Directors has arranged to donate the entirety of its public content to the Library of Congress to archive. (Public content means anything that is accessible to everyone.)

H23. Researchers should be able to explore the archive now.

H24. The general public should be able to explore the archive now.

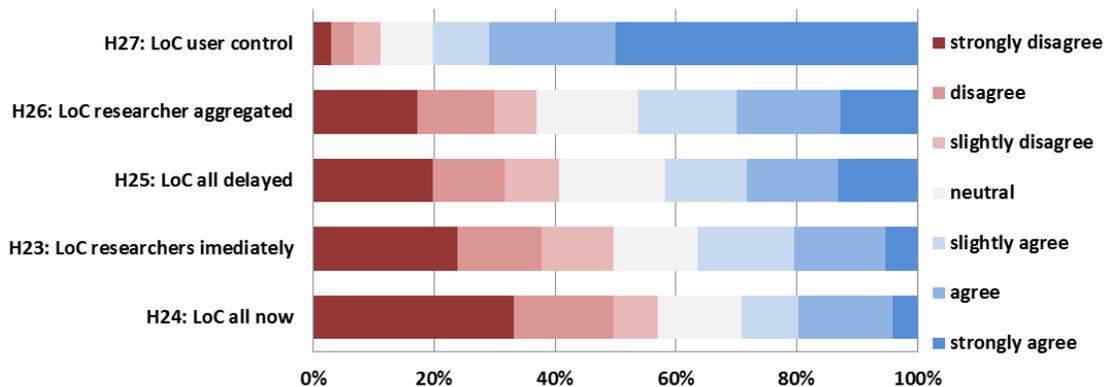
H25. The general public should be able to explore the archive in 50 years.

H26. Researchers should be able to run programs to explore aggregate data now

H27. Facebook should offer its users control of which material is archived.

**Figure 1. Facebook institutional archiving hypotheticals.**

Figure 2 shows the participant responses. The differences in responses are significant for all pairs except for the H25 (50 year embargo) and the H26 (researcher access to aggregated data



**Figure 2. Participant responses to the 5 conditions shown in Figure 1.**

only);  $p < .001$  (Wilcoxon signed rank test). There was little objection to archiving when users have control over what is archived. But such control imposes new burdens on users and begs the question of when users will give this permission: when they set up the account (before they know what they will use it for); after they're done using the account, when they are unlikely to expend further effort; when the institution is constructing the archive (which is unlikely to be well-aligned with the users' activities). Limiting access to aggregated data (which presumably anonymizes much of the content) and placing embargoes on access are the next best solutions although an archive with such access policies is still highly controversial to our respondents.

The scenario exposed participants to the ideas of embargoing the information and limiting access to the archive to researchers as potential remedies to privacy concerns. This question also gave participants the ability to express any saved-up horror at concepts introduced in earlier scenarios (e.g. monetization or transfer of the service to a new owner). Reactions varied considerably, from a sense of inevitability ("It is on the internet already so they are free to use it anyway." [SN014]) to outrage ("I think it would feel like a huge violation of privacy." [SN096]). Similarly, participants were divided about the archive's utility. To some it fulfilled imperatives ("...I'm a historian by training and I think there will be a wealth of interesting information there in 40 years..." [SN099]); to others, there was really no point in saving what was intrinsically ephemeral ("...It's not really public information. Imagine the garbage that teenagers and young 20's post. Do we really need to have that in the public records? No. Facebook is not a serious manner of communicating." [SN215]). Many participants felt strongly about their answers. Despite the fact that this question was at the end of a long survey, some participants wrote lengthy responses explaining their reasoning.

We used the following coding rules to tabulate participants' answers: If the response began *No*, and explained why, we coded it as a *no* even if the explanation implied that there might be conditions under which archiving was okay (e.g. with explicit per-item permission). If it began with a conditional (*if* or *only if*), we coded it as a *conditional yes*. If they began *Yes*, and were comfortable with the idea of archiving social network data, or believed they could do nothing to stop it, we coded the answer as a *yes*. Table 6 shows these results.

**Table 6. Should Facebook be archived by a public institution?**

Response	#	%
No	128	52%
Yes	54	22%
Yes, conditionally	57	23%
I'm not sure	5	2%
<b>Total</b>	<b>244</b>	<b>100%</b>

We saw many distinct reasons for objecting to or supporting the hypothetical institutional archiving initiative. For example, at least five participants realized that Facebook's terms and conditions may allow the content to be donated to and archived by a public institution. Of these five, only one said this meant that the archiving could proceed (*I would not like it if a company archived the information but I believe they have the right to do it as long as Facebook has it within their terms of service.* [SN157]); three others felt that the effort would not hold up, given broader ethical concerns, e.g. SN173 said, *I understand the legal aspects and that all FB users agree to FB's terms of use, but I feel that we live in a culture in which people do not take these terms seriously, and this*

*allows FB and other entities to abuse their customers information. While this may all be legal, I do not believe it is ethical.*

### 3.2.1 Objections

Let's look at the reasoning behind participants' objections to archiving Facebook content. They follow four lines of argument:

- (1) Archiving violates privacy and ownership principles;
- (2) Malicious applications may outweigh beneficial ones;
- (3) Content veracity and timeliness cannot be guaranteed; and
- (4) There is no societal benefit to archiving Facebook.

**Loss of control.** Participants who are opposed to a prospective institutional archiving effort generally have a keen sense of how it might violate their right to control the material they post. Almost a quarter of participants cited their right to privacy as a primary reason a public institution should not be able to archive Facebook; e.g., SN012 (one of the more thoughtful privacy advocates) said, *I guess it is a little hypocritical for me to [object]. I don't want anybody looking through my stuff. When I post on Facebook, it's for myself and my family and friends, not for the whole world. People said that if you post anything on the Internet, it's for everyone, but when people use social networking they believe the contents posted is only for their contacts, not for everyone...* In other words, SN012 thought that content would be exposed beyond the scope of its intended audience. Others simply cited a right to privacy. SN207 added that it's the persistence of content that is troubling *...most of the time people don't post things they'd like to remember forever ...*

Other participants invoked a notion of content ownership and copyright protection, asserting that archiving represents copyright infringement, or that it enables future infringement through reuse. E.g. SN132 said, *This stuff does not belong to them. They have no right to it. This is the work process of those who contribute to facebook.* SN193 more strongly asserted ownership: *It's MY data and I feel that I'm entitled to the rights of it, no one else. It should be my choice what content and with whom it is to be shared.*

There is a sense that permission might mitigate these rights violations (although the remedy was insufficient to allow archiving to proceed). e.g. SN150 said, *[Archiving should not go forward] without permission. Average citizens are not public figures. There is already too much of an erosion of our constitutional rights and it is perpetuated by government and business.*

**Projected inappropriate future use.** Some participants based their judgment on projected future use of the archive's content; they speculated that malicious use of the personal content would outweigh beneficial applications. Some of the future uses participants foresaw were overtly malicious (e.g. identity theft) or surveillance-oriented; others could simply not identify a legitimate use, and based their arguments on a sense that the content would be used strictly to satisfy nosiness (e.g. SN196 said *I don't want strangers all up in my business*). Others pinned inappropriate future use on the ability to search (e.g. *I wouldn't want people searching through my personal account.* [SN077]) or envisioned more targeted violations: *...Much of my information online is compartmentalized (restricted from certain people's view) for reasons of my own safety. I have abusive parents, and am a lesbian who is heavily involved in the queer community of my town, and it would be dangerous for people like my employers (I do not live in a state that has adopted the Employment Non-Discrimination Act) to see that I'm queer.* [SN174]

Several participants read us the riot act for even considering this option: even research use or plain curiosity might violate social norms. SN147 (female, born in the 1980s) said that, *I don't post information for it to be archived and studied. I do it for myself. Even though the Internet is very large, public, and anonymous, so is the real world. It'd be like a trusted friend spilling your biggest secret. Everything doesn't have to be known just because the information is out there.* Similarly SN180 said, *whether it is public or not, [institutions] really should not have a right to [archive personal content]. What ever happened to people being "nosey"? If you do not have consent, you do not have a right. That's all there really is to it and if you disagree, then you're in the minority and need to get a clue. ...*

**Content veracity and timeliness.** Past studies have identified peoples' concern with the correctness and timeliness of online information [16]. Since Facebook profiles and posts are proxies for the people they represent, certainly participants might be concerned about their accuracy. e.g. ... *Facebook profiles may not accurately represent a persons [sic] character, credibility, moral values, etc.* [SN204]). More importantly, people worried that they would be unable to reinvent themselves. For example, SN003 fretted, *the person i was years ago is not the same person today* and SN087 generalized, *it isn't fair to do that to people. People change and some people change A LOT and archiving things that people do and using that to judge them is very unfair.*

**No societal benefit.** A smaller subset of participants objected to Facebook archiving on the basis of its ultimate value to society (e.g. SN101 said, *...I don't think items on Facebook are important enough to be stored for future generations*). SN126 (and several others) held up the Twitter archive as comparable: *...Why is the Library of Congress archiving tweets... What is the point of keeping a hundred billion justin beiber tweets?*

A few participants did not understand the notion of public institutions, and how they compared to private corporations like Facebook. This misunderstanding underscores the need to explain this kind of effort. For example, after asserting that it was not okay for a public institution to archive Facebook, SN140 said, *maybe the government [should be able to archive it] but no privately owned company should have access to Facebook's data.*

### 3.2.2 Assent

Table 6 shows that between a quarter and a fifth of the participants accepted the idea of a Facebook archive; a relatively small number were openly enthusiastic about it. The arguments followed four lines, from principled to wholly pragmatic:

- (1) Archiving Facebook is a project for social good;
- (2) Facebook content is in the public domain anyway;
- (3) Facebook content is ultimately harmless and trivial; or
- (4) We are powerless to prevent this, so why worry about it.

**Social good.** Although people are wary of institutional archiving efforts [16] and have a wavering sense of the value of Facebook content [12, 5], a few people saw value in the content and some way around the privacy concerns. Only two participants perceived the project to be inherently worthy (*...Facebook is a social institution and deserves to be preserved.* [SN024]) and a third thought the content was potentially valuable in an ahistorical way.

**Public domain.** Some participants see Facebook data the same way they see all online content: If it's online, it's in the public domain. SN015 said as much, and several others qualified similar statements (e.g. ... *It should be understood that anything posted*

*on a site like Facebook is effectively being put into the public domain, so if a person does not want that information to be out there then they shouldn't be posting it in the first place. I think a lot of people have unrealistic expectations on this issue.* [SN209]) Others were less convinced that their content was in the public domain, but they did not question a public institution's ability to access and store it.

**Ultimately harmless or trivial.** While assertions that Facebook content is ultimately harmless or trivial are not powerful support for mandating archival attention, they do suggest that a segment of Facebook users would not actively obstruct this effort (e.g. ... *there is nothing there to be concerned about. I don't share a lot or make controversial postings.* [SN127])

**Powerless to stop it or it's up to Facebook.** Finally, it is possible to sense an attitude of futility. Some participants felt their opinions were meaningless in spite of a mild horror over the inevitable loss of privacy: they had brought it on themselves. SN168 said, *I feel that it would be a personal invasion of privacy. Although I feel this way, I understand that I chose to use Facebook, and I chose to put my information on there. Basically, I would feel invaded, but it would be my own fault because Facebook is a public service.*

### 3.2.3 Conditions

Although we only counted the positive comments in Table 6, many of the negative comments expressed a desire to impose some constraints if archiving were to go forward. Participants elaborated on the constraints we suggested in earlier portions of the survey to raise archiving and access conditions including:

- (1) Only public portions should be archived;
- (2) Content owners must grant permission or opt-in;
- (3) Embargoes that delay use should be imposed; and
- (4) Access and use restrictions should be implemented.

The first two conditions are by far the most common; at least 30 participants discussed the need for permission to harvest the content and 35 specified that only public portions of Facebook should be included in the archive. The second two conditions are less common: 6 participants mentioned content embargoes, and another 6 cited researcher-only access. Other use-based restrictions appeared in ones and twos.

**Only public.** Participants have a variety of ideas about what constitutes the public part of Facebook: Does this mean public/fan pages? Does this imply construction of an archive that contains material controlled by each Facebook account's privacy settings? Does it mean adhering to public/private divisions by communication types: Are messages all automatically excluded because they are by nature private communication? (e.g. ...*Public content is like a literary work, and literary works are stored.* [SN200]; ...*If it's publicly posted, it's public and people shouldn't gripe about where it's going.* [SN165]; ...*Because anyone can archive any public information they want to now, there's really no difference between an institution or a person doing it* [SN074]; *It would only be okay to archive the things that are already public. I don't want my profile and postings available to the general public, that's why it is private now.* [SN109])

Nonetheless, the spirit of these responses is much the same: public content may be subject to archiving imperatives and private content should be exempted. That so many participants brought up the "only public content" constraint suggests it is an important matter to discuss; however, it should be noted that archives of

personal material are built on foundations of letters and other forms of correspondence. Adhering to this constraint will change the nature and utility of the archive. An archive that is groomed to be strictly public information might ensure an anodyne source of historical information, less informative than a local newspaper.

**With permission or opt-in.** Permission is a popular concept in this survey, as well as in past surveys about institutional archiving and reuse [16]. Requesting individuals' permission or allowing each person to designate what goes into an archive and what does not is a conservative approach to constructing this sort of resource. Generally participants interpret this to mean that the permission is granted when the archive is constructed, not as the material is reused. SN191 elaborated that he believed an archive should include *ONLY the parts which I deem okay to archive. It should be an opt-in option only*. A few others imposed more general systems of permission that might apply to content by type, and by categories of reuse, e.g., *I think the site should specify what type of information will be archived, when they will be access [sic] and who should be able to access them. They should also give us the choices of what can be archived from our profiles*. [SN218]

**Embargoes.** One way of reducing an archive's privacy concerns is to delay access, usually by what might be considered a lifetime. Currently the US Census, a useful resource for biographers, historians, demographers, and genealogists, among others, is delayed for 72 years. Other resources (such as the Social Security Death Index) are constrained by their nature: records are released after the subject dies. This type of embargo arose in a small number of responses. SN053, who did not support the idea of a Facebook archive, said: *... I might be game if it was 50 years from now. By then it's a quaint cultural artifact with no real barrings [sic] on anything by then*. SN110 supported the idea of the archive, but said he'd support it, *only if the data was accessed long after I was dead. Say in 75 years*.

**Use and access restrictions.** Of course, it is not the archive itself that poses threats to privacy or to copyright holders; rather it is access. Thus one way to address these threats is to control access by restricting it to users with known motives (e.g. demographers) or to activities deemed appropriate (e.g. research). As we would expect from the objections, people are generally afraid of copyright violations or privacy breaches: *...It's only troublesome thinking about who gets their hands on it after it becomes archived. So if I post some nice photography that I've done and it becomes archived and then some stock photo company takes it and starts selling it, I start to have a problem with that. For research and historical purposes I don't really mind, but people can't be trusted to behave themselves*. [SN165]

Finally, participants reacted to controversial recent news about whether Internet-scale data sources can be used by law enforcement or surveillance agencies. Those who mentioned this type of use seemed divided on whether it explicitly could be used this way in the public interest (*...I could see if the "public institution" was private and had some kind of reason behind it, i.e. they think illegal activities might be happening...* [SN183]) or yet another way privacy would lost (*[archives could be constructed from Facebook data] as long as it wasn't sold to employers or used against people in some way* [SN052]).

We are left with the impression that Facebook's own user population is at best unenthusiastic about the idea of institutional efforts to archive the content of the social networking service. They offer compelling arguments for relegating this data to the dustbin of history: a desire to maintain privacy, the apparent

fluidity (and volatility) of the resource, a lack of any guarantee of veracity, potential violations of intellectual property, and most importantly, a perceived lack of long-term value. It is this unconquerable trump card that we address in our subsequent discussion: why archive Facebook if no one cares what's in it?

## 4. An Exercise in Historical Biography

To motivate further discussion, we discuss recent biographical research undertaken by the first author; we bring this research into the picture for two reasons: (1) to show how different types of data sources can be brought together to support inquiry; (2) to explore the apparent disappearance of these data sources in a modern age, and to show how Facebook has taken their place.

William S. Burroughs has been the subject of considerable literary interest over the past four decades. Originally associated with the Beat Movement in the 1950s and early 1960s, Mr. Burroughs has gone on to become a postmodern literary celebrity. He has been dubbed the godfather of Punk, and has been reinterpreted under the rubric of Queer Theory [23]. Dozens of literary biographies have been written about him, including Barry Miles' recent comprehensive effort, *Call Me Burroughs* [19].

Most Beat biographers devote at least some attention to Joan Vollmer, Burroughs' common-law wife from the mid-1940s until he accidentally shot her to death in Mexico City in 1951. It is not only the salacious nature of her fate, but also her apparent influence on writers such as Burroughs, Kerouac, and Ginsberg, that has kept her in the biographical limelight. However, until recently, she was seldom a research subject on her own, in part because of the paucity of information about her life: her death occurred before any of these writers were famous. Hence valuable historical materials and ephemera such as letters and personal papers were discarded or lost. By this time, most eyewitnesses are dead (and it might be argued they were unreliable to begin with).

However, since Burroughs' death in 1997, several investigations have been conducted, centering on Vollmer's short, tragic life. In 2002, James Grauerholz, Burroughs' personal secretary published an important monograph investigating Vollmer's death [6]. Rather than being a straightforward exculpatory exercise, Grauerholz instead brought together and examined as much evidence as he could to allow the reader to decide whether the murder was intentional. This effort opened the door for further scholarship. In 2007, University of Texas Brownsville student Christopher Carmona re-examined her life in his Master's thesis. His primary contribution was to request her academic records from Barnard College, where she enrolled as an undergraduate in fall 1939. Scholars and biographers had always taken it as a given that she had graduated from Barnard. It was there, while she was living in an apartment near Columbia University, that she met William S. Burroughs, Allen Ginsberg, and Jack Kerouac, and was credited as influential in shaping the group's philosophy.

Based on these records from Barnard, Carmona asserted that Vollmer had not graduated after all. Instead, she had been dropped from her classes in January of 1940 without completing a single term after she'd eloped with "a Mr. Adams". Her considerable literary influence took place without the benefit of a traditional undergraduate education. [3]

### 4.1 Retracing Joan Vollmer's steps

In November 2013, this paper's first author became interested in extending Joan Vollmer's biography; the research that follows represents a small portion of what she has found, and is intended

as a provocation about the unforeseen value of ephemera and news sources that are akin to what we now record in Facebook.

In the time since Carmona’s research, the embargo on the 1940 US Census expired. If Ms. Vollmer left Barnard in January, 1940 to elope with Mr. Adams, and she was still married to him when her daughter was born in 1944, they would likely be living together when the census was taken in April. But census records showed no Paul and Joan Adams in New York City (nor elsewhere in the US), even with relaxed age constraints.

A digitized collection of New York state newspapers, which included the *Albany Evening Journal* and the *Knickerbocker News*, revealed multiple references to Ms. Vollmer as she was growing up: as a Girl Scout; as a high school student; and as the winner of a \$500 scholarship to attend Barnard—these are all consistent with biographies. More surprising, however, was a brief announcement that appeared in the *Knickerbocker News*’s society pages on January 20<sup>th</sup>, 1940: On January 9<sup>th</sup>, Joan Vollmer wed Mr. Henry Allen Keeler of Great Neck, Long Island in Henderson, North Carolina. How had this announcement appeared without comment?

When a 50-year embargo on Henderson marriage licenses expires, they can be requested for research or genealogy. The license document returned from a print request to Henderson’s Office of Vital Records tells an intriguing story. First, Ms. Vollmer reported her age as 18 (she was still 16 in January 1940). The Henderson City Directory identifies the witnesses as a town doctor, a hotel manager, and the officiating Episcopal minister’s wife; it also gives us the addresses for the train station, the doctor’s practice, Henderson city hall, the church, and the hotel. A map reveals that the first four addresses are within a block of each other, and the hotel is a short way down the street. Switching to a street view, one can see that the church’s steeple is visible from the train station. A railroad map corroborates that Henderson is the first town over the North Carolina state line (the town’s website refers to Henderson as The Gateway City).

But who is Mr. Keeler? His obituary, which appeared in the *Jefferson City Post Tribune* in February, 1978, tells us that he attended Columbia University. Digital archives of the *Spectator*, Columbia’s student newspaper, reveal that Henry “Hal” Keeler was a freshman in 1939, the same year Ms. Vollmer enrolled at Barnard. He cultivated an active social life: he joined a fraternity and ran for freshman class president in those early months of his first year. The *Spectator* and Barnard’s student newspaper, the *Bulletin*, describe a whirl of dances held so the incoming freshmen at the two schools would mix and meet.

So they met and quickly eloped. But what happened next? The *Bulletin* indicated that finals didn’t begin until late January. Could Ms. Vollmer, now Mrs. Keeler, have slipped back into school and finished her first term? Special collections librarians at Barnard helped establish that she appeared in both the 1940 and 1941 yearbooks, hinting that she had continued her schooling. But this does not square with Carmona’s description of the records he had requested from Barnard’s registrar. An initial enquiry confirmed that she was dropped from her classes before finals.

Virginia Gildersleeve served as Dean at Barnard from 1911-1947. Through articles in the *Bulletin*, and Celine Young’s letters to Allen Ginsberg (Young was Vollmer’s Barnard peer and a member of the same social circle), Dean Gildersleeve’s flinty temperament is revealed. In a letter to Allen Ginsberg, Ms Young says of her request to sit in on Lionel Trilling’s class at Columbia:

“No, we will not be sharing Mr. Trilling’s class together, should Mr. Trilling permit notwithstanding. Dean Gildersleeve put both feet down and raised both eyebrows – an agile feat for an elderly dean, and so we shall abide academically on both sides of Broadway, much as I shall miss the company of Mr. Trilling and Mr. Ginsberg.” [from Ginsberg correspondence, Stanford University]

Thus it is no surprise that a memo dated January 16, 1940 from Dean Gildersleeve specifies that Joan Vollmer should be considered “dropped” from Barnard because:

“She ran away and got married after Christmas, (January 9th) having first made a false statement regarding where she was going, with deliberate intent to deceive. This falsification of the record is a very serious matter, and I do not think Ms. Vollmer is entitled to honorable dismissal.” [1]

At the memo’s end, Dean Gildersleeve added that Barnard might consider taking Vollmer back in the future. If we consult the 1940 census armed with the knowledge that Miss Vollmer would likely be Joan Keeler in April of 1940 (when the census was performed), and that the Keeler family lived in Great Neck, Long Island, we find a record for Mrs. Joan Keeler, 17, living with her new husband and in-laws.

What happened to this ill-advised first marriage? World War II Army enlistment records show that when Mr. Keeler joined the US Army in October of 1942, he reported his status as single. Nor did his obituary mention a first wife; he remarried in 1943. Joan Keeler reappears in the 1941 Barnard yearbook as Joan Vollmer. The hundreds of newspaper reports of her death in 1951 never refer to this first marriage. It is likely the marriage was annulled.

This youthful indiscretion gives us some insight into Ms. Vollmer’s life prior to the published accounts. That Henry Keeler was a “square” by Beat standards (by way of confirmation, he later graduated from law school and became a small town judge in Missouri) tells us a different story than so many Beat texts tell us about the disaffected young woman who is fed up with the Western literary canon and disillusioned by conventional thought. [20] Instead, she seems to have been caught up by events at hand and a more conventional desire for independence. That she repeated the same scenario in 1942 to become one of many “war brides” confirms this; it suggests that she had a mercurial side, willing to change course quickly and sharply. It is no wonder that by 1945, after her first child and her involvement with Burroughs, her peers consider her wise beyond her years. Both Kerouac and Burroughs refer to her as “older” than others in their social circle [19]; in point of fact, she is a year younger than Kerouac.

	ephemeral	recorded/archival
embargoed	correspondence (print, microfilm)	US census data (digital) US WWII military records (digital) small town licenses (print)
public	city directories registers maps (digital) photos	student newspapers (digital) small town newspapers (digital) yearbooks (print) climate data

Figure 3. An overview of data sources

## 4.2 Sources for biographical research

Let’s look at the sources that went into this biographical tidbit. Figure 3 divides them by type, whether they were publicly available at the time, or their availability was delayed, and whether they were regarded as ephemera or kept as a matter of record. We categorize maps as ephemera because map imagery is

refreshed frequently. Each type of data has analogs in Facebook, and can inform a sensible policy of what to do with Facebook's web of data.

#### 4.2.1 Embargoed data

The data associated with Joan Vollmer's story was subject to multiple embargoes: As we have already mentioned, census data is embargoed as a privacy protection measure. Each state or federal census collects different details, but generally involves names, addresses, ages, household relationships, occupations, housing costs, income, and other personal information. Detailed military enlistment records are similarly protected. Other data was embargoed pending the death of those who might be harmed by its release. Lucien Carr, a Beat ringleader, was convicted of the 1944 murder of David Kammerer (Burroughs' childhood friend). Data associated with the period, including personal papers and a narrative account of the murder, *The Hippos Were Boiled in their Tanks*, were embargoed until Carr's death in 2005.

#### 4.2.2 Ephemera

Public and personal ephemera were useful in reconstructing Joan Vollmer's life during this time period. A 1939-1940 city directory for Henderson, North Carolina not only specified personal and business names, numbers, and addresses, it also gave a secondary source for professions, and went through the city street-by-street. This directory helped construct the story behind the elopement and convey a sense of what Henderson was like in 1940.

#### 4.2.3 Private data

Many of the special collections associated with this circle of people contained private correspondence; members of the Beat Generation were inveterate letter-writers, and published volumes of their letters stand alongside their other writings. These letters were not released to researchers until their recipients—in this case, Ginsberg and Parker—were ready to donate their papers to university archives. This research took advantage in particular of the Allen Ginsberg Collection at Stanford University, although the collection antedated Joan Vollmer's first marriage: the events described in Section 4.1 took place before Vollmer met Ginsberg.

Dean Gildersleeve's memo clarified Joan Vollmer's dismissal from Barnard in 1940. It explains why she did not complete her first term of college, and poses a possibility that she could have returned to Barnard subsequently. (Unfortunately, while we can learn something from the presence of a particular record, we cannot infer the opposite from its absence.)

## 5. Conclusion

Now we can speculate about the role of Facebook in an imagined future project analogous to the one we describe in Section 4. Archiving Facebook will doubtlessly be a balancing act. On one hand, people have legitimate concerns about the threat to privacy and loss of control of one's own digital footprint that such a collection represents. On the other hand, a Facebook archive may provide a rich historical resource that documents a time period in the way that fast-disappearing resources like local newspapers, city directories, and personal correspondence did in the past.

To achieve this balance will be difficult. Hence we shouldn't just ask if a Facebook archive is valuable—despite our own past doubts, now we believe it will be—we should also ask, valuable to whom? Valuable when? Valuable how? At the present time, in many places in the world, Facebook is the social network of record (even among communities where it has been rejected in favor of newer services, it is still grudgingly used). In future work,

we will explore additional data about how participants feel about the possibility of Facebook data being monetized, and used by different commercial concerns, or of Facebook being replaced by another major social network.

Survey results suggest, however, that the following questions be resolved in the name of balancing competing interests of historical value, personal privacy, and accurate self-representation:

(1) What kind of institution is sufficiently trusted and competent to not only create this archive, but also to store it and mete out access as policy dictates? Should the institution be governmental? This is a controversial question: scandals such as recent revelations about the NSA's data gathering practices might make us look to an international non-governmental body to do this sort of archiving. But sustainability must be guaranteed, and the organization must have the technical wherewithal and trust to be able to collect and store the data and avoid data breaches.

(2) What is the nature of down-the-road data privacy as it relates to Facebook content? Participants are privacy sensitive, generally with good reason. But the privacy is a complicated matter, and Facebook data and privacy controls may be used for a variety of different reasons. It is different for someone to keep a relationship status protected from a prying aunt than for someone to keep their email address out of the hands of spammers, or to keep a message private because it is highly personal or a post private because an employer might read it. Certainly it is easy to see the potential for damage if this data is not protected. Furthermore, people have different understandings of personal privacy and different abilities to implement their own privacy standards. Some private data is marked as private, and other private data relies on privacy-through-obscurity to keep it from the public eye [8]. Although there have been thorough studies of privacy in the moment, it would be useful to think about privacy from a genuinely historical vantage point, outside the realm of its value to the individual. Easy mechanisms to opt-in to data donation plans may help, so people can donate correspondence in the way that they currently donate materials to special collections.

(3) How should digital data be embargoed and for how long? Unlike the Internet Archive's copies of major portions of the public Web, a Facebook archive may need to be embargoed in much the same way that census data (which is similarly personal) is embargoed to assuage peoples' concerns. A basic embargo period for Facebook will need to be computed using actuarial data. During this period, any archival snapshots of the data must be kept private; presumably the "keeping current" aspect of personal curation is an important aspect of privacy. This is in fact demonstrated by our survey's removal results. Digital data has a way of leaking through easily-made copies. Even physical data, once it is digital is very easy to copy, store, and share. For example, a special collection, which may be an important and valuable institutional asset, can leak via a series of high resolution cell phone photos as researchers form their own sub-collections. Hence to archive Facebook, the data will need to be encrypted, which may complicate—but should not stand in the way of—normal digital archiving and curation practices.

(4) How should we handle (and potentially record) change? We see from our study and others [12] that profiles are kept up to date rather than simply kept. People actively destroy social network content in the name of keeping things current or taking a break [22]. This question interacts with our call to understand privacy: it is clear from the study results, and from past studies, that one reason people worry about an archived copy is that it will no

longer reflect who they are now or their ability to reinvent themselves. Yet this data is useful from a historical perspective.

(5) Who may access the data, and how may it be used? Is a full embargo necessary for every form of the data and every user of the data? The US Census (which is mandated by the US Constitution) is released in several different ways [27]. As we have discussed, a full copy of the census is available once the embargo period has completed. In the interim, some of the data is available to people like journalists, with the constraint that if it is used inappropriately, the journalist loses access to the resource. Aggregate data is available before the full embargo is lifted; the greater the degree of aggregation, the earlier the release. Need also dictates the release of some of the data (i.e. for political redistricting). Hence we would advocate a tiered model of data release according to the content's sensitivity, the degree to which the data has been aggregated or otherwise anonymized, the need demonstrated by the data's user(s), and consonant with big data privacy issues [2]. The qualitative results shown in Figure 2 suggest tiered release will be less controversial than more liberal access policies.

We might have originally interpreted the results of our study as arguing mildly against archiving Facebook; a Facebook archive garners minimal popular support, and a number of Facebook users argue that the data they are creating is trivial. Instead, the biographical research we report here has changed our perspective: Temporal Facebook snapshots that balance privacy concerns with the desire to maintain a rich record of this era will create a valuable resource for future researchers. Such a resource must be handled appropriately, both to guarantee that it is useful in the way we expect it to be, and to address the valid concerns expressed by our study participants. Requirements for a large-scale endeavor like this one should come from at least two sources: potential users of the resource in the future and from social norms surrounding the resource as it is used today.

## 6. Acknowledgments

We would like to thank Yoram Meroz for lots of strategy help with the census and other resources, Special Collections Librarians Sabrina Sondhi (Columbia) and Tim Noakes (Stanford), archivists Jocelyn K. Wilk (Columbia) and Martha Tenney (Barnard), Director of Archives Shannon O'Neill (Barnard), and Constance Brown (Registrar, Barnard) for their help untangling Ms. Vollmer's complex and often misstated biography. This work was supported in part by National Science Foundation grant DUE-0938074.

## 7. REFERENCES

1. Bernard Registrar, private correspondence to the first author dated 5 March 2014.
2. boyd, d. and Crawford, K. Six Provocations for Big Data. Oxford Internet Institute, 21 September 2011.
3. Carmona, C. *The Legacy of Joan Vollmer: The Short Significant Life of a Beat Woman*. Master's Thesis, University of Texas, Brownsville, 2007.
4. Farnham, S. and Churchill, E. Faceted identity, faceted lives: Social and technical issues with being yourself online. *Proc. CSCW 2011*, pp. 359–368.
5. Good, K. D. From scrapbook to Facebook: A history of personal media assemblage and archives. *New Media & Society* 15, 4 (2013), pp. 557–573.
6. Grauerholz, J.W. The Death of Joan Vollmer Burroughs: What Really Happened? Fifth Congress of the Americas. Universidad de las Americas, Puebla, MX, 18 Oct 2001.
7. Gulotta, R., Odom, W., Forlizzi, J. and Faste, H. Digital artifacts as legacy: Exploring the lifespan and value of digital data. *Proc. CHI 2013*, pp. 1813–1822.
8. Herzog, W. & Stutzman, F. The Case for Online Obscurity, *California Law Review*, Vol. 101.
9. John, J. L., Rowlands, I., Williams, P. and Dean, K. Digital lives: personal digital archives for the 21st century: an initial synthesis. *Version 0.2* (3 March 2010).
10. Karnik, M., Oakley, I., Venkatanathan, J., Spiliotopoulos, T. and Nisi, V. Uses & gratifications of a Facebook media sharing group. In *Proc. CSCW 2013*, pp. 821–826.
11. Kirk, D. and Sellen, A. On human remains: Values and practice in the home archiving of cherished objects. *ACM Trans. Comput.-Hum. Interact.* 17, 3 (2010), p. 10.
12. 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.
13. Marchionini, G. *Information Concepts: from Books to Cyberspace Identities*. Morgan & Claypool, 2011.
14. Marshall, C., McCown, F. and Nelson, M. L. Evaluating personal archiving strategies for internet-based information. In *Proc. IS&T Archiving 2007*, pp. 151–56.
15. Marshall, C.C. and Lindley, S. Motivations and Strategies for Self-Search. *Proc. CHI'14*, pp. 3675–3684.
16. Marshall, C.C., and Shipman, F.M. On the Institutional Archiving of Social Media. *Proc. JCDL'12*, pp. 1–10.
17. Marshall, C.C., and Shipman, F.M. Experiences Surveying the Crowd: Reflections on Methods, Participation, and Reliability, *Proc. WebSci'13*, pp. 234–243.
18. McCown, F., Nelson, M.L. and Van de Sompel, H. 2009. Everyone is a curator: human-assisted preservation for ore aggregations. In *Proc. DigCCurr 2009*, arXiv:0901.4571.
19. Miles, Barry *Call Me Burroughs*. Ingram, 2014
20. Morgan, Ted *Literary Outlaw: The Life and Times of William S. Burroughs*. Norton, 2012.
21. Odom, W. Sellen, A., Harper, R. and Thereska, E. Lost in translation: Understanding the possession of digital things in the cloud. In *Proc. CHI 2012*, pp. 781–790.
22. Rainie, L., Smith, A. and Duggan, M. (2013). Coming and going on Facebook. *Pew Research Center's Internet and American Life Project*.
23. Russell, Jamie. *Queer Burroughs*. Palgrave Macmillan, 2001.
24. Shipman, F.M. and Marshall, C.C. Creating and sharing records of multiplayer online game play: practices and attitudes, *Proc. ICWSM'14*, pp. 456–465.
25. Stutzman F. and Kramer-Duffield, J. Friends only: examining a privacy-enhancing behavior in facebook. *Proc. CHI '10*, pp. 1553–1562.
26. Zhao, X. and Lindley, S. Curation through Use: Understanding the Personal Value of Social Media. *CHI'14*, pp. 2431–2440.
27. [http://www.census.gov/privacy/data\\_protection/](http://www.census.gov/privacy/data_protection/)
28. <http://digitalpreservation.gov/personalarchiving/>
29. <http://www.statisticbrain.com/facebook-statistics>