Rethinking the Web as a Personal Archive

Siân Lindley*, Catherine C. Marshall[†], Richard Banks*, Abigail Sellen*, Tim Regan*

*Microsoft Research Cambridge
21 Station Road, Cambridge, CB1 2FB, UK

[†]Microsoft Research Silicon Valley
1065 La Avenida, Mountain View, CA, 94093, US

{sianl, cathymar, rbanks, asellen, timregan} @microsoft.com

ABSTRACT

In recent years the Web has evolved substantially, transforming from a place where we primarily find information to a place where we also leave, share and keep it. This presents a fresh set of challenges for the management of personal information, which include how to underpin greater awareness and more control over digital belongings and other personally meaningful content that is hosted online. In the study reported here, we follow up on research that suggests a sense of ownership and control can be reinforced by federating online content as a virtual, single store; we do this by conducting interviews with 14 individuals about their Web-based content. Participants were asked to give the researchers a tour of online content that is personally meaningful to them; to perform a search for themselves in order to uncover additional content; and to respond to a series of design envisionments. We examine whether there is any value in an integrated personal archive that would automatically update and serve firstly, as a source of information regarding the content within it (e.g. where it is stored, who has the rights to it), and secondly, as a resource for crafting personal artefacts such as scrapbooks, CVs and gifts for others. Our analysis leads us to reject the concept of a single archive. Instead, we present a framework of five different types of online content, each of which has separate implications for personal information management.

Categories and Subject Descriptors

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

Keywords

Personal information management; virtual possession; ownership; control; Cloud; social media.

1. INTRODUCTION

In recent years, the Web has become a nexus for sharing, publishing, and storing personal content. It is a place where most of us leave a substantive digital footprint, through deliberate design as we publish media, but also indirectly, through communicative activities and by sharing content within our social networks. In other words, whether we mean it to or not, the Web has increasingly become a place where much of our personal content is archived. Indeed, recent features like Facebook's *Timeline* acknowledge the potential value of archiving new forms of content, such as social media. Such shifts raise the question of what this user-contributed content means for the future of the Web and Web-based services. Moreover, what kinds of tools do people need to manage, maintain, keep, and keep track of the content that matters to them on the Web?

Consider, for example, one personal media type: photos. At first blush, the problem seems tractable: most of us share informal photos with friends and family on a few social media sites like Facebook and Twitter or via email. We may keep our more artistic photographic efforts on a site like Flickr to reach a different audience (other serious amateur photographers, perhaps). We may use a generalized Cloud storage service on the Web, say Dropbox, to back up photos from our phone's camera and to move the photos between devices before selecting the ones we want to share or publish. But there may also be other photos we care about. Someone may have tagged us in their own Facebook photos; these photos can disappear without warning. Furthermore, Web sites and services are not static: a service can be shut down or an account can go dormant; we can forget how to access it; or we can forget it even exists. We may keep some of our photos by reference, e.g. by URLs that lead to someone else's Dropbox, or as email attachments kept on temporary storage. Indeed, loss on the Web has been attributed to many sources other than technology failure [8][13] and a person's sense of ownership may extend beyond her own photos [14].

Recent research has indicated that the lack of certainty as to *where* content that is hosted online really is has consequences for one's sense of control and ownership over it [18]. Yet, research has pointed to the value of underpinning this. Work with teenagers [19] and new university students [1] has illustrated how virtual artefacts such as Facebook photos and chat logs are printed off and displayed on bedroom walls or pasted into scrapbooks. Other work has found people may have alternative strategies, such as downloading content onto laptops and other local devices to create a stronger sense of ownership [18]. However, these approaches mean that potentially interesting digital attributes and social metadata, such as comments, tags, and likes, may be either lost or fixed in time. Giving users the ability to feel in control of online content, whilst at the same time being able to retain some of its valuable digital attributes, is an area worth exploring [6].

One way to meet this challenge is to provide tools that recognise that users may want more awareness of and control over their online digital belongings. Various researchers have suggested that ownership and control might be reinforced by federating content as a virtual, single store, although their suggestions differ in terms of how personal content ought to be pulled together and managed, and what it will be used for once it has been created. On the one hand, Odom et al. [18] propose an archive be constructed by unifying content in a virtual way from many distributed online sources, so that it can be viewed and managed as a whole. Alternatively, Marshall [11] suggests such an archive be created by federating metadata records in a centralized store until the actual online venues disappear, so that the content can be automatically saved at that time. The first solution assumes on-going interaction and crafting; the second, benign neglect. But both solutions raise a deeper question of federation: does it make sense to bring dispersed online personal resources back together as an archive?

In the research we report here, we sought to address this question first and foremost, and examine whether there is any value in an

integrated personal archive that would automatically update and serve as a source of information regarding the content within it (e.g. where it is stored, who has the rights to it). Relatedly, we explore whether this might then serve as a resource for crafting personal artefacts, such as scrapbooks, CVs and gifts for others. Such a tool is predicated on the assumption that people are interested in Web content that is about them, and that they might wish to find, keep, and use it to create new digital forms.

We explore these questions through interviews with 14 individuals, who were asked to give researchers a tour of personally meaningful online content, to search for themselves on the Web to uncover extra content that they might not be aware of or had forgotten about, and finally, to respond to a series of sketches, showing potential ways in which this content might be viewed together and managed, as well as serving as the basis for additional creative activities.

2. RELATED WORK

As indicated above, our study was motivated by an overarching question: is it valuable to be able to view and manage dispersed online resources together? We chose to address this by exploring the concept of an integrated personal archive, a notion that finds some support from previous work. Research on archiving has investigated both physical and digital content, and many researchers have tried to address both within the same studies. However, while studies of cherished physical materials have revealed a rich array of archiving practices, studies of digital content can, in contrast, highlight how easily the content can become lost or forgotten. Part of this contrast seems to hinge on the different ways in which digital and physical content can be stored and how they are then encountered. For example, Kirk and Sellen [10] have argued that the archiving of cherished objects entails their being enmeshed in the material fabric of the home. They separate materials that are displayed, e.g. a photo; objects that are used, e.g. a grandmother's ladle; and objects that are stored away, e.g. the family china. Together, these materials support a range of values, including defining the self (e.g. by displaying a particular photo), connecting with the past (e.g. by using the ladle), and fulfilling duty (e.g. by keeping the china safe, even if it is not to your taste).

However, this distinction between *safeguarding*, *using* and *displaying* physical mementos does not really extend to digital content. Digital things tend not to support functional storage, and while they might be put on display (e.g. a favourite photo might be selected as a screensaver), this often involves assigning a specific action to a file rather than changing one's understanding of where it is stored. Instead, digital file storage tends to mix the special and the mundane, prioritizing characteristics such as media type (e.g. having a folder for all photos), or relying on the default locations specified by software (and this is increasingly so, given the shift to operating systems which render the file hierarchy invisible [6]). Having been filed away in this manner, digital content is infrequently sorted, distilled or even encountered [20], and while this neglect is usually unintentional [11], it nevertheless results in users finding it surprisingly difficult to pinpoint their valued digital content. People struggle to specify notable digital mementos [21], and even when they do remember a specific artefact (such as a favourite photo), they can find it difficult to locate, becoming frustrated and concerned in the process [27]. As noted though, benign neglect can be less risky than ill-conceived personal information management strategies. For example, the British Library's Digital Lives program [8] and also Marshall, McCown and Nelson [13] have highlighted the problems of storing digital possessions in the Cloud. These are often lost, simply

because accounts are forgotten or become impossible to access, or because policies are misunderstood. Furthermore, research has suggested that users have a weakened sense of possession over content that is stored in the Cloud [18]; the ambiguity regarding *where* precisely it is stored somehow undermines this.

The research described so far is predicated on the familiar notion of standalone artefacts that can be managed, such as photos that can be filed away in a folder, or uploaded to the Cloud. However, personal digital content is increasingly taking forms that move away from this fundamental concept. Indeed, in the same way that material things become enmeshed in the fabric of the home [10], we might consider personal digital things as becoming increasingly integrated in the fabric of the Web. For example, Harper et al. [6] point to social media as things that one might wish to download or otherwise act upon, but that do not support the simple range of actions normally associated with files. One cannot, for example, simply *save* a status update as a standalone object, or *copy* a photo that integrates the social metadata that is associated with it. Harper et al. suggest new actions are needed, which better enable users to act upon, and thus feel in control of, their online content. They suggest that such actions are essential if users are to have a greater sense of control, and ability to manage, digital content in a socially-networked world. Furthermore, sharing via social media raises questions regarding who might have the right to save or copy content. People feel that they can save and reuse photos taken and posted by others [14], and recent studies have indicated that users (particularly teenagers and young adults) download and print out social media, sometimes along with its metadata (such as comments), for display or to be incorporated into scrapbooks [19][1]. We might expect this to be especially true for personally relevant social media, such as photos that one is tagged in; tagging might be interpreted as a way of extending ownership [1].

In this paper, we take a personal archiving stance toward federating disparate content sources. We focus on content that might be kept over time, abstracted from its immediate use, rather than a personal information management (PIM) perspective that focuses on how this content may be managed in the context of daily activities. Yet there is much to be learned from a PIM approach, because it harbours many of the same assumptions as a unified archive does: the need for a uniform, user-defined, underlying schema [9]; that fragmentation can be conquered using a linked object-based approach [25]; that search may be more important than structure for eventual access [3][4]; and that use can pivot around a central organizing paradigm such as email [2]. We suggest that this literature be kept in mind as a resource for design once the basic concepts underlying a unified archive have been tested, as we do in the remainder of this paper.

3. METHOD

3.1 Participants

We interviewed 14 individuals, 8 in the UK and 6 in the USA. We specifically recruited people who we expected to have a substantial online presence (e.g. musicians, serious amateur photographers, bloggers, and gamers), in addition to users of widely-adopted social networking services such as Facebook. We were careful to include two digital natives amongst our participants, whom we might expect to demonstrate different keeping and sharing practices than other cohorts (see e.g. [19]). See Table 1 for full details.

Participant	Age	Gender	Based	Details
Harry	35	M	UK	Amateur musician, consult-ant
Charlie	36	M	UK	Amateur musician, academ-ic
Jacob	27	M	UK	Amateur musician, sound designer
Oliver	48	M	UK	Amateur photographer, academic
Ava	38	F	UK	Blogger/tweeter, e-learning professional
Thomas	32	M	UK	Blogger/tweeter, lobbyist
Jane	32	F	UK	Stay-at-home mum
Sophie	18	F	UK	Student embarking on a gap year
Ann	50	F	US	Tech writer, amateur musi-cian
Kim	50	F	US	Novelist, blogger and stay-at-home mom
Lynn	37	F	US	Admin, videographer
Todd	33	M	US	Journalist, fact-checker
Mary	31	F	US	Non-profit administrator
Vincent	20	M	US	Student, math nerd

Table 1. Details of the participants. Pseudonyms are used to protect anonymity throughout the paper.

3.2 Interviews

Interviews were semi-structured and open-ended. They comprised three parts: an online tour, a self-search exercise, and reactions to design envisionments. For the online tour, participants showed us how they created and stored digital content, and the sites and services that constituted their online presence. Prompts were helpful in eliciting this information (for example, we asked whether they used Facebook, how they shared photos, whether they had ever looked for a classmate, whether they blogged, and so on). We also asked participants to search for themselves, with the idea that we would not only turn up additional content they had forgotten, but also that the search would reveal content that was *about* them, digital belongings outside of their control, but potentially interesting or valuable to them. This search would also uncover doppelgangers, provoking participants to discuss what was theirs and what emphatically wasn't. The tour and self-search exercise culminated in the participant sketching out important components of their online lives and how these fit together (see Figure 1).

The design envisionments took the form of 15 sketches (all can be seen at [24]). Three are simple ways of presenting content that is dispersed across the Web in the guise of a single, integrated archive, for example, presenting content on Facebook, Flickr and Twitter together. Six show different ways of viewing this content, by filtering it (e.g. by where it is located, who posted it, or who has rights to view it) and organising it as a timeline. The remaining six show different artefacts that might be crafted out of this content, representing a diary, photo album (Figure 2), scrapbook, greetings card, resume, and a gamer-oriented memento.

The interviews lasted from 90 minutes to almost three hours, depending on the scope and diversity of participants' online presence. Participants were given a £40/$65 gift voucher to thank them for their time. The interviews were audio-recorded and, where relevant, photos were taken and screen captures saved. All interviews were then transcribed and the transcripts were analysed for emergent themes. These are presented in the following section.
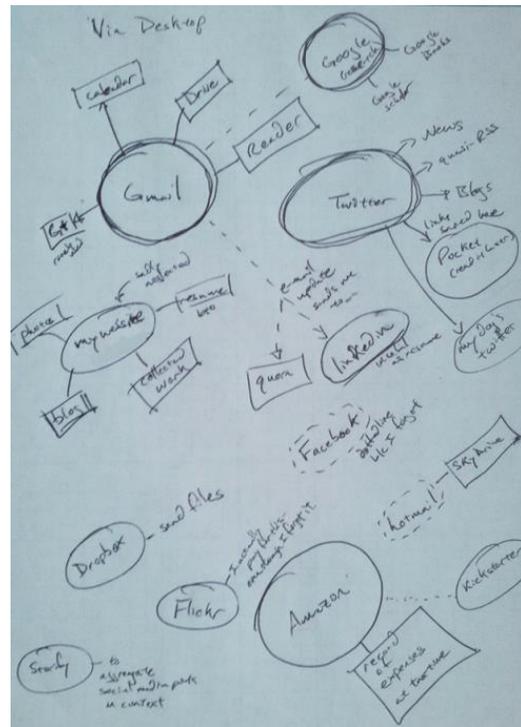

Figure 1. Todd's sketch of his online content.

Figure 2a. Envisionment sketch of a diary (journal), with annotated events drawn from online media activity.

Figure 2b. Envisionment sketch of a photo album crafted out of various online media.

4. FINDINGS

The findings are organised into two main sections. While the interview procedure had three parts, many of the findings that emerged during the second part (self-search) echoed those found during the first (online tour) and were linked thematically during our analysis. We present these findings together. Reactions to the design sketches, which were created to test our ideas about the

automatic construction of an integrated archive and the crafting activities that it might enable, are presented separately. In the following section, we focus on the online content that was uncovered during the online tour and through self-search, and reveal how uploading content and supporting its re-finding was part of a broader approach to managing digital possessions online. This treatment of the Web as a relatively permanent site for storage was nevertheless juxtaposed with evidence that participants led a trajectory across online services and social media accounts, embracing and then abandoning them with remarkable consistency.

4.1 Personally Meaningful Content Online

The online tour was designed to explore the kinds of things that people care about online: what they feel is about them, what they believe they own, what they would like to keep, and what they feel is dispensable. For example, did participants rely on Facebook (or a similar site) as the central nexus of their online lives? What happened to content outside of their immediate control? Did they handle content they created differently to that which they actively curated? How were obsolete profiles handled? The activity of self-searching often arose naturally as part of the online tours, and when this did not occur we asked participants to search for themselves once the online tour was complete. Although this tended to produce content similar to that already discussed, we asked participants to elaborate and refine their search terms, or tailor them for different profiles (including pseudonyms), with the aim of finding as much content as possible.

In what follows, we explore two overarching themes that emerged from the data: the Web as a reliable location that was incorporated into personal information management, and the trajectories that participants had experienced in moving across different online services and user accounts.

4.1.1 Personal Information Management Online

One of the main observations to come out of the online tour was that the uploading of user-generated content, as well as the curation of content generated by others, was part of a wider process of personal information management. This type of behaviour is known to be undertaken by a large proportion of Web users; recent statistics show that 46% of US adults who are online upload their own photos and videos to the internet, while 41% curate photos and videos they find there [22]. Here we examine some of the reasons driving, and values supported by, these activities.

4.1.1.1 Websites as Archival

Most straightforwardly, our participants considered some of the content that they showed us to have been archived on the Web. These archives might encompass content that solely existed online (e.g. emails in sites like Gmail), content that was backed up to an online location (Todd, who saw his tweets as a significant part of his professional online presence, backed them up to his Word-Press blog) or content that was explicitly uploaded (e.g. photos on sites like Flickr). This reflected something of a shift from local storage, with Ava noting during her interview, "*I am not really talking about my actual devices in terms of file storage at all now, which is quite weird*". Typically, these websites had certain archival qualities. Both Flickr and Gmail were noted for making content "*searchable*" (Oliver), and sites such as Gmail are perfectly suited to the benign neglect that users often demonstrate when it comes to managing their content. Some participants, such as Jane, confessed to never deleting any emails and relying entirely on Gmail's search feature to manage her email. Todd, who no longer accessed his Pro Flickr account, nevertheless continued to pay for the storage because Flickr's keywords and geo-tags made it "*important to me in terms of finding stuff and retrieving things*".

Interestingly, the process of uploading content to the Web, and of deciding where to put it, was integral to some of the decision-making that underpins personal information management. Oliver "*triaged*" his photos, with only "*the ones that I like*" going to Flickr, and Ava, who also had a Pro Flickr account, uploaded only "*good photos*" from her "*proper camera*" to the site. For both, the process of uploading to Flickr was part of a wider process of digital archiving, which entailed deciding which images were worth keeping, and which were not. Interestingly, this decision was also sometimes bound up with *which* sites participants uploaded content to. Ava noted that photos that were "*just capturing a moment out with friends*" were *not* backed up, although they *were* posted to Facebook. In other words, despite being stored online in much the same way as a Flickr collection, content posted to Facebook was not seen as having been secured. This may be partly related to the quality of the online versions, but for Ava, there was also a conceptual distinction between the two sites. One site was archival, for things she wanted to keep, the other ephemeral, for things she wanted to share in the moment.

In addition to being part of the process of deciding which content is worth keeping, uploading content to the internet was also a way of making certain it was accessible to specific others. Amongst our US participants especially, we found instances where couples shared a content repository, usually of photos or videos. The content nominally belonged to one person, but could be readily accessed by the other. For example, Mary and her boyfriend Jim share a Picasa account, held under his name, where they keep a joint collection of more than 20,000 photos and videos. Mary thinks of it as her most important online resource, subsuming what she might post in Facebook or Pinterest: "*everything lands in this Picasa bucket. Because all of my important videos and photos are in here ... If anything crashed I would hope it was not [Picasa].*" She stays "*permanently logged in at home*". Similarly, Kim and her husband Paul share a photo server as a source for online photo-sharing and publishing projects. He is the photographer and curator but they both dip into the photos for different projects. According to Kim, "*So this is how we do it. He has all these photos here, because he just takes gobs and gobs of photos. So he puts them on here partly so we can access them. Because he also puts them on Shutterfly. But if he puts them on Shutterfly, I can't get them off for free and put them on my website.*" Thus, storing content online is not only a means of keeping it for oneself. Personal information management practices are intertwined across individuals, and the Web plays an enabling role in this.

4.1.1.2 Finding as a Proxy for Keeping

In addition to websites that are recognised as repositories for particular types of content, such as photos or emails, participants included websites in their online tours that enabled them to keep track of content that had been uploaded by others. Sometimes this content was closely linked to the participant. For example, Sophie was tagged in 3,063 photos on Facebook, comprising those she had uploaded herself and those added by friends, the lines of ownership being blurred between the two. Similarly, friends of Charlie had uploaded videos of some of his gigs, tagging them with his Facebook profile: "*I think there's four bits of live footage, [by] people at the gig who we knew .. I've got links to them*". Being able to re-find this content was seen as something of a proxy for ownership; "having links" to photos and videos was described in similar terms to "having" the content itself.

This equating of re-finding and keeping was also extended to content that had no relation to the participant other than they had chosen to collect it. Ava described using a mixture of online re-

sources to curate online content, enabling her to make it re-findable: "*my personal blog, this was set up so that I could keep all e-learning things that I found of interest in one place .. now I don't need to do that because it is actually easier to put it onto Scoop.it, or even if I know I have seen something on Twitter and I favourite it or re-tweet it, I know I can find it again*". Our research also provides examples of how the boundaries between what is created and curated can blur. For example, Mary used Pinterest to organize a mix of content she had found online with her own pictures: "*It is things that I'm planning. I've definitely used Pinterest for my wedding more than anything else. So I'm planning a style and theme and what I want.*" Mary's collection seamlessly combines her own user-generated content with that generated by others to make a new and unique artefact.

In most cases, participants did not back up this type of content. This was especially the case for that uploaded by friends and family; here it was believed that the original could be sourced from whoever had originally uploaded it, should the need arise. Obviously, such reasoning could not be applied if the content had been uploaded by a stranger or if it had been generated online (e.g. as tweets or blog posts), and in exceptional cases, participants did attempt to archive this type of content themselves. For example, Vincent backed up YouTube videos and academic papers that he would be unable to recover if they were removed from the Web. In general though, participants did not give much thought to the vulnerability of re-finding as a strategy for managing content generated by others. The realisation of this, triggered by the interview, led Mary to exclaim, "*You're stressing me out! I need to back up my Pinterest. It just dawned on me, what if Pinterest died? Where would my wedding go? ... I don't know why I think Pinterest will live forever.*"

4.1.1.3 Keeping the Whole

Mary's realisation points to the complexity of backing up this type of content. She does not wish to back up the component parts of her Pinterest collection; she wants to back up her whole "wedding". Todd provided two further examples that illustrate the difficulty of backing up content that is entwined with a website. The first related to a desire to be able to download his dog's Facebook page, a site which had seen lively social activity as he, his wife, and their friends interacted and shared photos, but which was due to be removed by Facebook. The second was associated with his earlier appearance as a contestant on a TV game show, the details of which had been documented on a website maintained by obsessive fans. In both cases, Todd had contemplated ways of keeping the content, although none were satisfying. Speaking about the latter, he noted, "*the problem is, it's hard to know what's the way to save it ... If I wanted to preserve the actual page, as a functioning thing in the way that it was intended to look and feel, I just haven't thought about that*".

4.1.1.4 Summary

To sum up, some digital content is valuable in and of itself. Other content is valuable as part of a collection, or as a more complex digital entity. The value of user-generated content that can be manipulated as standalone artefacts, such as photos, can be reinforced through their being archived or shared on certain sites (e.g. where posting to Flickr reinforces the status of "good photos"), and this additionally serves as a means of backup. However, much of the content that only exists online is not backed up. Sometimes this reflects a belief that it can easily be re-found; sometimes it is not clear how one could back it up in a meaningful way.

4.1.2 Trajectories across Sites

That so few of our participants made efforts to back up their online content leads us to one of the contradictions in our data. In general, participants treated the Web as a means to archive or otherwise reliably re-find their content. Yet most also had a history of embracing and then abandoning particular websites and user accounts. As already noted, Todd's Pro Flickr account was something that he "*used to be really engaged by*", yet he almost forgot it when sketching out his content because he no longer actively used it. Such narratives tally with recent statistics: 20% of online adults in the US who do not currently use Facebook say they once did [23]. In this section, we explore some of the reasons why old accounts are abandoned in favour of new ones, and the consequences of this trajectory for digital possessions.

4.1.2.1 Reasons for Abandonment

A number of reasons for abandonment were evident in the data. Straightforwardly, turning points such as the need to pay for additional storage (such as when Charlie reached the limits of his free SoundCloud account), or even to maintain a for-pay account (such as Charlie's Pro Flickr account) could trigger a shift to an alternative site or signal the decline of a previously much-used one. But such triggers cannot offer a full explanation for why participants moved from one website to another. Indeed, some participants, such as Todd, continued to pay for accounts that they did not use.

Thus, we focus on two further reasons why participants moved on from sites they had previously favoured. The first relates to who else uses the sites. Ann commented, "*I'm on Google Plus and so are many of my friends, but nobody uses it anymore. Nobody goes there anymore. But it's true there was sort of a sorting effect – it seems like the more quality people were there for a while*". Here, Ann situates her trajectory across social networks in terms of a broader social movement. A consideration of who else uses a site took on an additional degree of importance for participants who wished to build an audience. Jacob described how "*if I was to start a band now I wouldn't even think twice about putting a MySpace page up... no one will really go there, you can put as many records as you like on BandCamp, which is what bands are doing these days*". The second reason for moving on from certain sites is also reflected in Jacob's comment. Websites were noted as supporting certain features that were of the moment. He continued, "*they do this thing where you set [an album] up and you can pay what you like [to download music], which is very trendy at the moment, apparently*". Similarly, Mary said, "*I did get on Classmates for a brief period of time, but I found that Facebook sort of [displaced it], because you can search for your school*", and Ava described how she was "*moving away from blogging. Because you have these new sites like Scoop.it and Pinterest and you know, the kind of synaptic websites*". By offering new ways of curating content, or on-trend ways of distributing it, these sites could attract new users.

4.1.2.2 Consequences of Abandonment

Unsurprisingly, this movement from site to site led to loss of access to old accounts. Jane commented that she had started a blog once, and tried out Twitter, but "*I can't even really remember my passwords*", and Todd lamented the loss of his Friendster profile, viewing it as "*a weird relic of my first social network*". He valued the profile because it marked him out as an old-timer in the world of social networking, but he could no longer view it: "*I don't even know what email I used to sign up for it, or what my name is, or how to get to my profile. I'm guessing that it's just Friendster.com/username. I don't know. I think I tried to get into it.*" We also saw the converse of this situation, where participants believed

content to have disappeared, only to discover that it was still online. Harry was surprised to find that a search for his Flickr username, which he had long ceased paying for, produced the photos still associated with his account.

Harry's Flickr account had been associated with "*a phase of taking more pictures*", and Jacob made a similar observation regarding his prior participation in forums: "*I used to follow bands around on tour .. and be a strange sort of weird super fan, and then I'd go on their forums and post and sort of chat to other fans.*" However, while we might expect such sites to be viewed sentimentally as reminders of earlier interests or phases in life, this did not turn out to be the case. Harry was surprised but not curious to see his old Flickr account, and Jacob's reaction to a seven-year old band website primarily regarded how the site appeared, rather than the content it contained: "*this was 2005, so yeah this is how the internet used to look .. isn't it strange how websites have changed so much*". Charlie was similarly underwhelmed by a four-year old MySpace profile. He commented, "*we're coming up to doing a new EP now, so maybe we'll have some kind of rah rah rah launch for that [on MySpace], erm but what we're now thinking of doing is .. a vanity release, a double A side piece of vinyl, just you know just to have something to give to the grandkids*". The vinyl was perceived as worth keeping in the long-term, in a way that a MySpace profile simply was not.

While Charlie's MySpace profile had simply fallen out of use, other participants had edited their abandoned profiles to indicate a lack of currency. As a simple example, Harry's previous band had altered their MySpace profile to indicate that they had split up. More radically, Jacob had completely overhauled an old band's MySpace page. "*We had like I dunno 8,000 friends on MySpace, which was a great thing at the time. We've got 3,000 now cos I think we've just deleted loads and we've got 11 weird top friends who, no one knows who these people are .. and taken off all the music and replaced it with experimental stuff .. we've kind of mashed up our own song, but we don't sound anything like this [laughs]*". Rather than being kept as a sentimental relic, the site had been completely revamped, and the band had created a new site for their work on BandCamp.

Making it obvious that a profile is no longer in use means that complications associated with it being public can be avoided, and our analysis revealed how disused profiles could become a problematic element within a wider digital identity. This was perhaps made most obvious when Mary inspected her LinkedIn profile. She commented, "*I don't care for LinkedIn. But a long time ago I created it. ... And I don't keep it up to date. It says I still do [my previous business]. That was sold two years ago. But people continue to add me. And so, [she's reading from her profile] 'Viewers of this profile also viewed'. Oh, that's funny. They looked at me and they looked at [my current boss]. I don't know. See, and that's weird too. Now I see that – why were you looking at me and my boss? And who are you?*" The ambivalence, and even anxiety, that Mary portrays here is chiefly related to what she described as the "*call to action*" that LinkedIn represents. She is concerned that her disused profile will set false expectations or invite contact based on outdated information. For similar reasons, Jane had deleted the online dating profile through which she met her husband. Jane did, however, bemoan the fact that she could no longer access this profile: "*In fact that was really annoying, the fact that you can't go back to your profile, you can't find it again. Because when I got married I wanted to, in my speech, quote some things from [my husband]'s initial profile and I tried to go back and find what he'd written, but we'd both deleted our profiles.*" It is easy to see why Jane deleted her account; keeping an online dating profile

once in a relationship can clearly create a set of problematic social expectations. Once it was gone though, there was no way of recovering the content.

4.1.2.3 Reasons for Keeping
Given that abandoned social network profiles tended to be encountered without nostalgia, it is interesting to unpack more carefully why they were kept at all. For Mary, the social graph associated with her LinkedIn account made it difficult for her to delete it: "*The only reason I have it is because of the 'friends' function. I'm connected to a lot of – 89 people – if I take it down and I ever want it again later, well I'd have to rebuild it again. So I just keep it up because of the connections. But I don't like it. And I don't want it.*" Similarly, Ava saw value in her Twitter network rather than the tweets themselves, and said of her MySpace account, "*I don't think that it is precious in any way ... In fact I have logged in very, very recently but that was because I was trying to, in fact that is probably one of the reasons why you don't delete these things. My friend is getting married and I was trying to find his address somewhere*".

Yet being wedded to a site made some participants uneasy. Oliver described how he was taking steps to "*make me less dependent on the Flickr*", so as to avoid "*the entrapment of forever having to archive stuff there*". He had found a program that could download his Flickr archive and import it into iPhoto whilst retaining the tags, and had started the lengthy process of converting his entire archive. Similarly, Todd cited Twitter and Tumblr as potentially dangerous, because of the difficulty of pulling his content off of the sites.

4.1.2.4 Summary
User accounts are embraced and then abandoned as social networks migrate and fresh, innovative, features are introduced via new sites. Old profile pages may be kept, but these are a source of unease if public-facing because of the inaccurate message they convey. Consequently, profile pages are frequently hidden, deleted or revamped. If profiles are kept, it is typically for pragmatic reasons, such as access to a social graph, rather than nostalgia.

4.2 Responses to Sketches
In the final part of the interview, participants were asked to react to a series of design sketches. In this section, we examine some of the supporting assumptions, drawing on the (sometimes surprisingly adamant) responses of our participants. In particular, we highlight difficulties associated with the melding of distinct online identities, the long-term value of social media, and the applicability of paper-based genres (like scrapbooks and photo albums) to digital content.

4.2.1 A Single Archive Cannot Represent the Different Facets of Self
One of the main difficulties that participants expressed regarding the concept of an integrated archive related to this most fundamental aspect of it: its integration. The content that participants encountered during the online tour and self-search exercise was distributed across a variety of online repositories and social media profiles, and this distribution was meaningful. Content on different sites related to different facets of self, being associated with different online identities and intended for different audiences (see also [5]).

For example, Ava described her use of Pinterest as "*completely different from anything else that I do online … I don't even know if I would really like to engage either my friends or my professional contacts, because it is just really housewifey*". Most partic-

ipants had a professional identity that they wished to keep separate from other traces of their online behaviour. This was generally achieved through the use of different pseudonyms; for example, the musicians amongst our participants had various artists' profiles, which typically could not be connected to work-related identities. However, and as Ava's comment also indicates, this disconnect could not simply be viewed in terms of professional vs. non-professional. Various non-professional personas were also deemed to be distinct from personal identifies. Oliver said of his Flickr account, "*it would be impossible to link it to my name*", going on to comment, "*it's a very public private thing but, yes, it's just got nothing to do with me. You know?*" That a repository of 13,000 photos built up over many years has "*nothing to do with me*" seems a strong statement, but it was not unique. Thomas described his YouTube videos, which included content generated by, and featuring, himself, with the remarkably similar phrase: "*this isn't necessarily me*".

While the disconnect between different types of content is certainly related to self-presentation to others (participants relied on certain sites being visible only to specific audiences to preserve their privacy in an online, and very public, space), it is also a reflection of how participants understood themselves, and managed their own digital content. Mary knew that content to provide inspiration for her wedding was on Pinterest, while her outdated resume was hosted on LinkedIn. Presenting content from these two sites in the same folder then becomes problematic, and undoes much of the conceptual work that has already gone into managing it. Put simply, a central archive mixed things that simply weren't seen as belonging together.

4.2.2 Social Media Does Not Need to be Archived

Another problem with our envisioned integrated archive was that the type of content that dominated our sketches – Facebook updates, tweets, and other social media – was not seen as worth keeping by participants. Rather, this content was seen as ephemeral and not valuable in the long term. Jacob describes, "*you write something on Twitter and someone replies to it or blah blah blah, and then the next day it's kind of forgotten about ... I don't think people would want to go back to what they said*". This sentiment was common. Kim noted, while she was sketching out her online content, "*Then there's Facebook. I'll make it small... Huh. Well. I just can't think of anything in Facebook that I want to archive*", and Lynn, in spite of the fact that she reported devoting a considerable amount of time to the site, said, "*no, there's nothing on [Facebook] that I'd like to save*". In general, social media was seen as being for consumption by others rather than to be kept by oneself, and participants were generally of the view that it simply reinforced what they already knew about themselves. Jane said, "*looking back at my history isn't really so important because I can remember it*", and Harry agreed, "*it's not wildly interesting; I know who I am, I don't need to find stuff out about me*".

However, this viewpoint did become more muddled when it came to photos posted to social network sites. Photos as a media type were generally seen as the type of content that one should keep (participants often talked about wanting to keep photos, videos and music). Photos uploaded to Facebook were often described in different terms than those uploaded to sites such as Flickr (as we saw earlier), or kept on one's own disks. Thus even Sophie initially said of her 3,063 Facebook photos, "*I guess with Facebook they're really irrelevant stuff that you put up on there. So like, I love to cook, so people will take photos of what I've baked or something*". However, both she and Ava reconsidered this attitude during the interview, with Sophie reflecting, "*if I lost my photos on Facebook I'd be really upset, because I don't, then that's literally the photographic memory of my life gone for the last two and a half years*". Yet Ava could not imagine a practical way of reconciling Facebook content with an archiving strategy: "*So much of it is really mundane and the time that it would take to weed out anything important, to make that decision, so if you were going to back everything up, I mean how feasible is it really, realistically? You know, so much of it is just nonsense as well, I don't know, I really don't know.*"

As a final observation, the social metadata that was bound up with Facebook photos were also, generally speaking, not valued. Mary said, "*If my Facebook died, all those photos live somewhere else and I don't particularly care about the comments that much*". Similarly, Lynn speculated that she would shift to another medium if a Facebook interaction began to take on emotional heft: "*So I'm going to directly email or send a letter to somebody telling them that they're amazing... Saying, 'Oh Cathy, your hair looks crazy in that photo' is just dust in the wind.*"

4.2.3 An Archive Should Contain the Remarkable (and Remarked Upon)

The previous point brings us to a quality that participants believed archives should have: they should only contain content about the "*key events, like you'd have some of your wedding photos and some of your baby photos*" (Jane). This emphasis goes some way to explaining why only certain types of lost content were missed; the online dating profile that formed the basis for your first date with your husband, or your first social network, might fall under this category.

Interestingly, social media was seen differently when participants focused on remarkable events; special events were likely to be documented and commented upon, and these comments, tweets, etc., might be worth keeping. Ava described how "*when I posted my dissertation on Twitter and LinkedIn ... I searched for the title of my dissertation to see if anyone had blogged about it or anything like that*", and Kim used the title of her book to keep specific tabs on its post-release progress, and to turn up new reviews. As Charlie summarises, "*I already know that I played a gig here, I played a gig there, and I appear on this compilation and I appear on that compilation, so .. that's just noise really, what I'd like to know is what someone's said about it .. it would be good to be cross-referencing Twitter, maybe if you could delve into Facebook, just to see if anyone's talking about it, that's the only thing I'm really interested in, I know everything else*".

Thus, including information about the "*social life*" (Charlie) of a file, be this "*what people are saying about it*", the number of people who had consumed it or marked it as a favourite, or how it had been reused (this latter quality being especially relevant for photos) was of interest. In fact, Kim cited Google Analytics as one of the more important parts of her online presence: "*So [Google Analytics] is what I go to a lot to check how my blog is doing ... Partly because I'm curious. This will tell me what are the popular posts. Where the traffic is coming from... I did posts on our recent trip per city. I did one on Amsterdam, so that one seems to be popular. I'm getting a lot of hits on that one...*" Similarly, Charlie and Harry both noted that there was something interesting in knowing *who* was talking about your content. Charlie described how "*people you know are likely to be biased*" thus reactions from "*people you don't know .. shows something about the reach of the music*".

However, difficulties in drawing on content generated by others were also highlighted. Firstly, it is not necessarily the route to

finding content that is meaningful to users. As Harry noted, "*our most viewed or rated [Flickr photo] was a picture of somebody's arse*"; similarly, a co-opted picture of a nose that Kim had used in her blog "*has gotten more hits than almost anything else I've posted on my blog*". Secondly, there is a risk that it can challenge one's self-perception, and this is particularly problematic in the context of creating a personal archive. As Jane elaborated, "*everyone must feel that they've got their own self-image .. and I'd probably want my archive to reinforce what I already thought about myself. I probably wouldn't want any pictures in which I looked ugly, or any comments where people were dissing something that I'd said*". Similarly, Charlie commented, "*what I want to find, if I'm doing stuff on the kind of music profiles if you like, I'd like to find people saying nice things about me*". Thus photos should be flattering, reviews positive, and content significant.

4.2.4 Value May be Realised through Crafting

As participants were generally uninterested in the idea of archiving online content about themselves, we might expect the notion of using it as a resource in crafting to be similarly rebuffed. However, some participants did see value in being able to draw together their online media – especially media pertaining to key events in their lives – in new and expressive ways.

For this reason, the photo album sketch was one of the most popular of the design envisionments presented to participants. The focus on a single event (in the example, a birthday party) that was significant enough to be associated with media hosted across various sites, resonated with participants. Charlie described, "*if there's a gig there'll be usually different angles on it, actually something like this per event so, yeah so here's that gig you played in [this city] on that date, here's some photos which were taken, here's some videos, here's a recording of the gig, here's what some people said, I can see that that would be quite a nice aggregation*".

The potential to re-appropriate content was also noted in relation to this sketch. While Sophie had initially expressed uncertainty regarding keeping comments associated with Facebook photos, she noted the potential to instead use tweets as photo captions in the context of crafting: "*if you did a specific event it would be nice to have all the detail and like what people were saying about it ... if you have the photo and then you have like, someone tweeted 'So-And-So's wasted – hashtag'.*" She emphasized that, while tweets in themselves weren't worth keeping, "*if you had it as a caption next to a photo of someone who tweeted it, that would be nice*". This sentiment was echoed by US participant Ann, who said: "*That would be a handy way of – like if you had a bunch of crap relating to an event – and you wanted to revisit it and possibly pull out things from it ... And the little bits of things [i.e. comments] would probably be helpful.*"

Other sketches were well-received by individual participants for a range of idiosyncratic reasons. For example, Jane saw a value in creating an archive for her baby son; Mary saw the potential of collaging for family members; Jacob, who had recently graduated, spoke of his need for a digital portfolio; both Lynn ("*I can see last time I ran was the 18th*") and Vincent ("*I have these thoughts. I type them... And you have all the things you worked on before that. And I need to do these things in the future. It helps you organize yourself*") saw the mixed-media journal as potentially helpful from a life-hacker perspective. However, while the journal's pragmatic aspect held sway with Vincent, he was the most sceptical about the emotional resonance of the crafted material. After the interviewer finished describing the sketches, he became exasperated:

"*Who the hell looks at picture albums anymore? ... Put it like this. This entire thing [the crafting designs] has a fake picture album aesthetic. Why would you have a fake picture album? Except if you were nostalgic for picture albums. We don't use them anymore. At all. We just look at people's pictures in their profiles. And that's it. Picture albums have ceased existing. You don't use them at all. So you can't suggest, why don't you make a picture album. Because why would you do that? You just look at the pictures! The metaphor has disappeared. People like us – people like me – have grown up in this [way].*"

5. Discussion

The analysis presented above highlights various ways in which the Web is used as part of a personal information management strategy, but it also reveals how, despite the investment that is made with regard to particular websites, it is not uncommon for these to be abandoned or overhauled, practices that inevitably lead to loss of content. Despite the potential for loss, however, participants rejected the notion of a single integrated archive. The concept mixed content that was, in their view, distinct and more usefully kept separate. In this discussion, we tease apart five types of content identified through our analysis, each of which offers different implications for design. We then use these distinctions to explore how the Web and Web services can better support the personal archival practices that are already informally in place.

5.1 Five Types of Content

The categorical distinctions we discuss hinge on three recurring themes we observed in the data:

- The user's curatorial intent: Is the collection shaped and controlled intentionally, or does it accumulate through use?
- The digital original's disposition: Is the digital original local or online, and is it fully under the user's control or not?
- The collection's dynamic nature: Does the collection change additively or are changes necessarily destructive?

The interplay between these themes helps to distinguish the types from one another and to work out their implications.

5.1.1 High Value Collections

Our first category refers to collections of content that are uploaded to an online site, either as a means of sharing one's best work (e.g. in photography) or to back up particular content. These websites then become an integral part of personal information management, both as storage and as a record of what has been identified as high value. Additionally, these websites may offer useful functionality, such as tagging. Users may initially use such features for social reasons, but doing so has the consequence that the online collection acquires a richer layer of metadata than the originals, which may become a useful archival feature.

Because these sites have been selected to host a selection of content, they are easily articulated as containing meaningful artefacts. This resonates with research on material mementos and cherished objects; having a location for special content is part of the work of archiving and reinforces its status as meaningful [10]. Thus, these key websites were contrasted with other sites that might host the same type of media (e.g. photos), but that were conceptually associated with trivia and the ephemeral. The locale at which content is hosted is one way in which users make sense of their online content, with different sites being perceived as separate and distinct in meaningful ways.

5.1.1.1 Implications

This content has two key characteristics: it is associated with a digital original that is also under the user's control, and it is curat-

ed, a process that is intrinsic to selecting and uploading it to the Web. As such, this content is understood as having been safeguarded, and users see little peril in leaving it where it is. Yet our data reveal that even high value collections can hardly be considered a locus of permanence. For example, three participants with significant Flickr collections had all but forgotten about their accounts (one was surprised to find that it still existed); a fourth was taking steps to avoid being "entrapped" by the site. This is not problematic in the short term, but of course the value of a backup lies in it being reliably available in the future. If users can lose track of content stored in accounts that are only a few years old, how likely is it they might remember it in 10, 20 or even 50 years? And how likely is it that their content will still be there?

For content that is backed up to the Web, a key implication for design seems to be to *make visible links between the digital original and online versions of the same content*. Such a scheme can take advantage of URIs or the unique identifiers many sites assign to the content at upload time and may benefit from the systems community's research on tracking item provenance [16]. A file system that informs the user that a photo is also hosted online, and that possibly incorporates some of the features of those websites (such as tags, comments, or favourites), would not only reinforce the archival qualities of those websites, but could also serve as a reminder that content is backed up, where to, and perhaps if that account is becoming vulnerable (e.g. through disuse).

5.1.2 Collections that are Curated Online
A second category refers to content that primarily exists and is curated online. Examples include Pinterest boards, blogs, and even collections of tweets. These collections are hosted online and largely comprise other-generated content. There is no digital original under the user's control, and while duplicates exist (e.g. the person who uploaded a photo that has been pinned on Pinterest may hold the original and may host it elsewhere online), these would be difficult to obtain by the user if their collection was lost (even if simply because it's hard to remember what's in the collection). Thus, the collection is a means of curating but also supporting the re-finding of content that has been encountered online.

5.1.2.1 Implications
Collections of other-generated content are vulnerable to loss, and so implications chiefly relate to supporting their backup. Services already exist to support the backup of Pinterest boards, tweets and blogs, and this can take advantage of the curation that has already occurred online. For example, paid-for services such as pin4ever.com enable users to also store the visual juxtapositions, re-pins and likes that relate to pinned content. Such sites enable users to *retain aspects of the site that lend that content meaning*, and also provide the advantage of *permitting a local backup*. This seems essential: as users transition across sites and services, they are only likely to pay for backup accounts while they are active users of the primary site. As active use falls off, users are unlikely to remember their once carefully curated collections, and still less likely to accidentally stumble across them. Having copies stored with other valued content makes future encounters with them more likely, even if they are largely forgotten.

5.1.3 Collections that Emerge through Use
A third category stems from collections that are not so much formed through curation as generated through use. Examples here include folders of webmail that evolve over time, and the social graphs that emerge via social network services. In both cases, the user has made many incremental decisions regarding whether or not to store an email or accept a friend request, and the result is a collection of content that gradually accumulates.

5.1.3.1 Implications
Content in this category is similar to collections that have been curated online, in that it is not related to a digital original or duplicate that the user can easily access, and that it is the result of a curation of sorts. However, rather than coherent collections, these were often seen as useful resources, to be kept just in case. For example, it was interesting to note that within social network sites, *the social graph was often deemed more important to keep than the content itself*. Implications for design include support for search (e.g. being able to re-find old emails or locate contacts), and for ways to back up a social network that will remain meaningful over time. As social networks in particular are often intrinsically tied to a site, backup requires a consideration of how to support a compelling experience of browsing a disused network, particularly after a proportion of it has moved on to a new site.

5.1.4 Content for Consumption in the Moment
A fourth category is user-generated content posted to social network sites, such as Facebook and Twitter, appearing in news streams and associated with consumption in the moment. While social media can become a component of curated collections, it was generally seen as a means of communicating, enacting social responsibilities, and presenting one's face to the world. Thus, and despite the fact that a good deal of personal content is generated via them, social network sites were not seen as holding meaningful records or artefacts, and participants reacted strongly against the notion that one might wish to download and keep this type of social media.

What is particularly interesting here is that this view of social media as ephemeral seemed to generalise to media types that were, in other circumstances, deemed important to keep. Photos and videos, which were repeatedly cited as the types of content that one ought to safeguard, somehow became viewed as more trivial when hosted on a social network site like Facebook. Participants placed little value in the metadata associated with these files (such as comments and tags), which might be considered as embellishing the original. Further, when participants had no access to the digital original, they believed friends and family, rather than a website, were the best route to it. It was only in the circumstance of the interview that some participants began to reconsider these assumptions: would they in fact be upset if they lost their Facebook photos? Could they really rely on friends or family to produce all the YouTube videos that they appeared in?

5.1.4.1 Implications
As this type of content is generally viewed as not worth keeping, or as being reliably stored as a digital original elsewhere, participants did not see the need to back it up; nor did they wish for it to be automatically archived for them. This resonates with prior work [12] that has shown that people would rather lose some of their content than keep, and be overwhelmed by, all of it, no matter how automatic the keeping is. Unpacking frivolous from worthwhile social media presents huge difficulties, and may indicate why our participants made broader distinctions between their online content. Viewing everything on Flickr as valued and everything on Facebook as inconsequential may be a one-dimensional way of managing one's personal information, but it seems likely that richer strategies quickly become intractable.

However, exploring the concept of a personal archive with participants did reveal some ways in which social media might be leveraged. Firstly, while participants reacted against the need to keep social media that they had generated themselves, they did express an interest in knowing what people were saying about them, especially if those comments related to some key event or artefact,

such as a gig, mp3 release or thesis. Thus, one opportunity would be to support users in *identifying and linking (positive) social media to a digital original* that is located elsewhere; enabling users to browse social media by pivoting around key artefacts.

Secondly, our exploration of crafting revealed that, *while social media are not valued as standalone artefacts, they might be creatively combined in new and interesting ways*. For example, the re-appropriation of tweets as photo captions could support a new set of uses for them, which is associated with longevity rather than consumption in the moment. This resonates with recent research that has demonstrated the incorporation of social media in collages and scrapbooks [19].

5.1.5 Dynamic Content: Profiles and Personal Pages

A final category is represented by profiles and (less frequently) other personal Web pages, examples of dynamic content where changes are more apt to be destructive, as there is no easy way to undo edits. Furthermore, change is commonplace, as without it the user runs the risk of presenting an obsolete or outdated face to online communities. Although we were surprised at the lack of sentimentality with which old profiles were viewed, we saw two instances in which participants regretted having deleted, or lost access to, old profiles, and a third in which a participant was seeking a compelling way to back up his dog's Facebook profile. As with all personal content, one of the difficulties associated with user profiles is being able to predict what value they might hold in the future. But keeping profile pages presents a unique set of challenges. They are not generally seen as standalone artefacts which can be kept offline, but leaving disused, or out-of-date, profiles online presents its own complexities. Put simply, profiles are resources for action in the present, thus when moving on from old social network sites, users delete, remove or overhaul them in order to avoid setting misplaced social expectations. Where this is difficult to achieve, anxiety can result.

5.1.5.1 Implications

Profiles and personal Web pages make an interesting category because they are constantly in flux; content is lost frequently because it is overwritten or deleted by the users themselves, and these losses may only be regretted much later on. This suggests the possibility for a time machine-like interaction; one that allows the user to go back and revisit old profiles, rather than take steps to specifically back these up. Similar functionality is offered by sites like the Way Back Machine [26], but a personalised take on this, which archives a user's profile at moments of creation and then radical overhaul, may play a role in personal information management. Nelson et al. offer a mechanism for recovering this type of personal content after the fact from what they refer to as the Web Infrastructure [17], and McCown et al. have designed a mechanism for crowdsourcing link reliability in this context [15].

5.2 The Web as a Personal Archive

Our analysis highlights why drawing content from the Web to form an integrated archive does not offer a good solution for personal information management. However, we can draw a number of conclusions from the above, which underpin how the Web is, in itself, an archive of sorts, and how this usage might be better supported.

Place matters. This relates not only to the placement of content in certain sites with the aim of reaching different audiences, but also to how the meanings that become associated with those sites then impact the ways in which the content in itself is perceived. Drawing on Harrison and Dourish [7], we might suggest that these websites acquire a recognisable and persistent social meaning, which shapes how they are understood as places for particular types of content. Content on different sites is thus associated with different core values, which determine what is seen as worth keeping and what is superfluous.

Placing matters. That content is deliberately placed somewhere, rather than simply building up, is also important. Pinterest and Facebook are both social media sites of sorts, but in our study they were viewed very differently. One was seen as a site for curated and valued collections, one as an uncontrolled accumulation of trivia. Interestingly, social network sites were often valued more for the (curated) network that they represented, than the content that they delivered.

Storage matters. With physical artefacts, people can put their content on display, keep it in functional storage for practical usage, or place it into deep storage where it can be safely hidden away [10]. With online content, this distinction is weakened. Content is often on display by default, with no compelling option to keep it privately. Better options for dealing with dormant profile pages and archive sites, including ways of securing access in the future, could help here.

A coherent whole matters. There was a general sense that the keepable comprises photos, videos and music – content that resonates with long-held notions of what matters, and that takes the form of standalone artefacts. However, the Web comprises many other content types that may be worth keeping, including collections, social graphs and profile pages. These things are difficult to keep as a whole, but ways of doing so may prove valuable.

The digital original matters. Finally, users may judge the value of online content in relation to a separate, digital original. If stored locally, these originals may offer new opportunities for exploring online content, such as social media or archived versions of the same content. If stored remotely, other backup solutions may not have been considered, but they may be valued post-hoc.

6. Conclusion

This exploration of possibilities for archiving online content has led us to question many of the assumptions that we held going into this research. While we expected users to be concerned that their personally meaningful content is under the control of others, or hosted in ambiguous locations in the Cloud, we found that the notion of drawing it together, in a secure and centralised archive, raised a separate set of issues. Corralling this content in an archive juxtaposed what we have unpacked as five different categories of online content, each of which raise different implications for design in relation to personal information management. Fundamental to our analysis is the importance of place. In contrast to prior work, which has positioned online storage as somewhat ambiguous and consequently problematic, we saw that users drew on their understanding of the site of storage in managing and giving meaning to their own content. Paradoxically though, we saw that the embracing of certain sites was characterised by transience. Users invest in key sites, archiving their content or interacting with their network through them, and then overhaul or abandon them. That these sites cannot be safely stored away, and instead at best degrade, at worst are deleted or forgotten, raises a fresh set of challenges. Resolving the paradox of how to keep online content, which is public-facing and carries expectations of being up to date, is key.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Bales, E. and Lindley, S. 2013. Supporting a sense of connectedness: meaningful things in the lives of new university students. In *Proc. CSCW 2013*, ACM, 1137-1146.

[2] Bellotti, V., Ducheneaut, N., Howard, M. A., Smith, I. E. and Grinter, R. E. Quality versus quantity: e-mail-centric task management and its relation with overload. *Human-Computer Interaction* 20, 1-2 (2005), 89-138.

[3] Cutrell, E., Dumais, S. T. and Teevan, J. Searching to eliminate personal information management. *Commun. ACM* 49, 1 (January 2006), 58-64.

[4] Dumais, S. T., Cutrell, E., Cadiz, J. J., Jancke, G., Sarin, R. and Robbins, D. C. 2003. Stuff I've Seen: a system for personal information retrieval and re-use. In *Proc. SIGIR 2003*, ACM, 72-79.

[5] Farnham, S. D. and Churchill, E. F. 2011. Faceted identity, faceted lives: social and technical issues with being yourself online. In *Proc. CSCW 2011*, ACM, 359-368.

[6] Harper, R. et al. 2013. What is a file? In *Proc. CSCW 2013*, ACM, 1125-1136.

[7] Harrison, S. and Dourish, P. 1996. Re-place-ing space: the roles of place and space in collaborative systems. In *Proc. CSCW 1996*, ACM, 67-76.

[8] 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* (March 3, 2010).

[9] Jones, W., Phuwanartnurak, A. J., Gill, R. and Bruce, H. 2005. Don't take my folders away! Organizing personal information to get things done. In *Proc. CHI 2005*, ACM, 1505-1508.

[10] 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) 1-43.

[11] Marshall, C. C. Rethinking personal digital archiving, part 2: implications for services, applications, and institutions. *D-Lib Magazine* 14, 3/4 (2008).

[12] Marshall, C. C., Bly, S. and Brun-Cottan, F. 2006. The long term fate of our personal digital belongings: toward a shared model for personal archives. In *Proc. Archiving 2006*, 25-30, arXiv:0704.3653.

[13] Marshall, C. C., McCown, F. and Nelson, M. L. 2007. Evaluating personal archiving strategies for internet-based information. In *Proc. Archiving 2007*, 151–56, arXiv:0704.3647.

[14] Marshall, C.C. and Shipman, F.M. 2011. The ownership and reuse of visual media. In *Proc. JCDL 2011*, ACM, 157-166.

[15] 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.

[16] Muniswamy-Reddy, K., Holland, D. A., Braun, U. and Seltzer, M. 2006. Provenance-aware storage systems. In *Proc. 2006 USENIX Annual Technical Conference*, 43-56.

[17] Nelson, M. L., McCown, F., Smith, J. A. and Klein, M. Using the Web infrastructure to preserve Web pages. *Int. J. Digit. Libr.* 6, 4 (2007), 327-349.

[18] Odom, W., Sellen, A., Harper, R. and Thereska, E. 2012. Lost in translation: understanding the possession of digital things in the Cloud. In *Proc. CHI 2012*, ACM, 781-790.

[19] Odom, W., Zimmerman, J. and Forlizzi, J. 2011. Teenagers and their virtual possessions: design opportunities and issues. In *Proc. CHI 2011*, ACM, 1491-1500.

[20] Petrelli, D. and Whittaker, S. Family memories in the home: contrasting physical and digital mementos. *Pers. Ubiquit. Comput.* 14, 2 (2010), 153-169.

[21] Petrelli, D., Whittaker, S. and Brockmeier, J. 2008. AutoTopography: what can physical mementos tell us about digital memories? In *Proc. CHI 2008*, ACM, 53-62.

[22] Rainie, L., Brenner, J. and Purcell, K. (2013). Photos and videos as social currency online. *Pew Research Center's Internet & American Life Project.* http://pewinternet.org/Reports/2012/Online-Pictures.aspx. Accessed 8th March, 2013.

[23] Rainie, L., Smith, A. and Duggan, M. (2013). Coming and going on Facebook. *Pew Research Center's Internet and American Life Project.* http://pewinternet.org/Reports/2013/Coming-and-going-on-facebook.aspx. Accessed 8[th t] March, 2013.

[24] The Web as a personal archive. http://research.microsoft.com/en-us/projects/webaspersonal archive/default.aspx. Accessed 8th March, 2013.

[25] Van Kleek, M., Moore, B., Karger, D., André, P. and schraefel, m. c. 2010. Atomate it! end-user context-sensitive automation using heterogeneous information sources on the Web. In *Proc. WWW 2010*, ACM, 951-960.

[26] Way Back Machine. http://archive.org/Web/Web.php. Accessed 8th March, 2013.

[27] Whittaker, S., Bergman, O. and Clough, P. Easy on that trigger dad: a study of long term family photo retrieval. *Pers. Ubiquit Comput.* 14, 1 (2010), 31-43.