The Long-Term Implications of the Increasing Loss of Control Over Our Personal Information

Max Van Kleek
Dept. of Computer Science
University of Oxford
max.van.kleek@cs.ox.ac.uk

Reuben Binns
Dept. of Computer Science
University of Oxford
reuben.binns@cs.ox.ac.uk

Jun Zhao
Dept. of Computer Science
University of Oxford
jun.zhao@cs.ox.ac.uk

Nigel Shadbolt
Dept. of Computer Science
University of Oxford
nigel.shadbolt@cs.ox.ac.uk

PIM2016: *For Richer, for Poorer, in Sickness or in Health... The Long-Term Management of Personal Information*. Workshop at CHI 2016, San Jose, CA.

http://pimworkshop.org/2016/

Abstract
Today, PIM practices are as shaped by the Web and cloud-based tools, much as the PC defined PIM activities nearly three decades ago. As personal information has transitioned to the many, popular, cloud-based apps, services, and platforms, however, people have unwittingly left the fate and future security of their personal information in the hands of a variety of commercial entities with a multitude of interests in the data at stake, essentially abdicating responsibility for its longitudinal preservation and continued accessibility. In this position paper, we discuss what we perceive to be the potential consequences of such choices will be, towards historically important PIM practices, and ways that people might once again re-gain control of their data in the future.

Author Keywords
Personal Information Management; long-term information needs; end-user control; privacy; Internet of Things; cloud computing

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

Introduction

Never have the economic forces driving the technology industry been more turbulent; each day marks the birth of a dozen services online, but also the death of another dozen. While today, a service used to keep track of routes through unfamiliar cities might go offline [2], tomorrow it might be something as fundamental as a Web identity provider [4] - a service necessary for people to get securely authenticated to any number of other online services. Caught in this sea are end-user individuals; ordinary people who signed up to services because they needed or wanted the services offered, and entrusted their data to them in exchange. Service that may be relied upon on a daily basis might be there one day, but gone, changed, or broken the next time it is needed.

In more competitive times, when a service shut down it was usually because the business behind it went bust. These days, despite the fact that information technology has become consolidated in the hands of just a few companies, not even these stable giants can be counted on to provide stable or long-lasting services [1]. New products are released to much fanfare, made available just long enough for users to become dependent upon them, only later to be discontinued without explanation or any means of graceful migration.

Such a situation fundamentally characterises perhaps the biggest PIM problem of our online lifestyle, resulting as a direct effect of business model of controlling people's data. This business model represents a fundamental shift from technology empowering people, to disenfranchising them *en masse* – eliminating people's ability to dictate how their own data are handled, shared, controlled, stored, accessed, and used. This control has shifted and is now centralising in the hands of a few, large, service providers.

Whilst we could delve at length into the history of the information economy to identify and explain the source of this shift, rooted in monetary incentives of making users dependent on services over the long term, we instead focus on the many implications of the loss of control within to the context of this workshop, "For Richer, for Poorer, in Sickness or in Health... The Long-Term Management of Personal Information". We briefly divide our concerns into five main areas: longitudinal preservation, collection, privacy, security, obsolescence, and trust, and discuss possible avenues by which control might be eventually restored.

Preservation

One of the most fundamental challenges with the long-term management of information is the ability to effectively preserve information for later use. While archiving digital information has always been a challenge for end-users even during simpler, "PC era" times, this has only gotten worse as people have abdicated responsibilities for information preservation to third parties in the "cloud". This is due to at least two reasons.

The first is that devices are going "cloud first". Everything from desktops to mobile phones, tablets and digital cameras, as well as the applications that run on them, are switching to storing information "in the cloud" by default instead of on the devices themselves. Specific examples range from all iOS devices, which sync photos, contact information, message, and so forth to iCloud, Android devices that do similarly, or Microsoft Office which syncs everything now to OneDrive.

This would not be a problem, except for the fundamental problems around user mental models of cloud services. Mental models for on-device storage are simple; in general people have no trouble understanding data stored in

a physical device, and that such physical devices will continue to “hold” the data indefinitely as long as the device is not erased, physically misplaced, or damaged. To mitigate potential loss, people can reasonably apply strategies they would use to ensure the security of a physical artefact to ensure these data remain available, such as making sure not to lose it or damage it. Moreover, the idea of making copies of the data on a device onto other physical devices is relatively easy to adopt.

With cloud storage, however, it is less clear to end-users that things they put in the cloud will last, if left untouched. Cloud providers exacerbate the problem by advertising the resilience of cloud storage, setting up false expectations that the data will remain available indefinitely. Unfortunately, data is no less volatile (arguably more so), and no more available (arguably less so), in cloud services, for more subtle reasons.

The very term ‘the cloud’ is misleading, since it gives the impression of a single unified place where all of the data belonging to one individual are stored. In fact, each individual has very many clouds, one for each service they use. And each one of these disparate clouds has its own means of organising its contents, making any kind of organisation across them inordinately difficult.

A second challenge to archiving is that one of the most basic abstractions for data since the dawn of computing is starting to disappear - the venerable file. Among the first of the many recent operating systems to do away with any user-visible notion of a filesystem was iOS, which eschewed notions of files by putting apps first, letting them handle data storage within them in bespoke, user-invisible ways.

There may be several reasons that files may seem like an antiquated approach to organising data. Files seem out of place in a world of structured microdata, where little pieces of data, whether they be tweets, likes, comments, instagrams and snapchats dominate the information landscape. Even previously discrete media items such as videos and audio files are now being blurred thanks to more sophisticated approaches to presenting them, sensitive to bandwidth, resolution, visual output modalities, interactive devices available, and so forth.

However, files serve a tremendously useful abstraction for end-user control, because, they provide a simple, robust model of information containment; a single, simple way of storing and organising data collections within hierarchies. Doing away with files eliminates people’s ability to think about reasonable ways to copy or back up their data, manage, manipulate and organise their data - how would I keep it if it wasn’t a file? Despite a rather illustrious and successful history of end-user structured databases like FIlemaker, such tools remain at the periphery of personal information management landscapes.

Creation of Personal Collections

Collecting is a fundamental activity that people have engaged in throughout the ages; music, books, films, stamps, photos, etc. Unfortunately, content networks have come to see personal collection as a violation of their property rights, because such collecting activities are seen to undermine their control of how and when copyrighted works are experienced by individuals. The move to the cloud is a means to further entrench such controls by requiring each experience of a copyrighted work to be approved by the content networks themselves. It thus threatens our ability to make ad-hoc collections of things we historically have been able to collect and experience in perpetuity.

Amazon's Kindle store and DRM-enabled reader devices and applications allow people to "purchase" e-books, but remain indefinitely in full control over the availability of whether, when, and how these ebooks can be read. Netflix lets end-users add films to their favourites list, but ultimately remains in control of how long these films remain available, and in which geographic regions. Music streaming services, such as Spotify, Apple Music, Grooveshark and others remain in control over which, when, and what audio is available to be streamed on their protected media applications and devices at all times.

Such dominant control by media companies, and networks has fundamentally even shaped the kinds of digital devices available on the marketplace. Five years ago, removable media personal video recording hardware were still readily available on the market, including DVD Recorders, and VCRs. As analogue TV was phased out, the transition to HDTV silently introduced much more effective mechanisms of control in the form of "broadcast flags" that allowed networks to fully control exactly how, why and how long recorded programmes could be kept and re-watched. In to be certified compatible with receiving HDTV signals, such broadcast flags had to be honoured with DRM software running on all recording devices. Since removable media devices allowed end-users to physically remove storage media, and therefore prevent devices from automatically deleting stored programmes, such removable media devices were silently deprecated and replaced with the now-pervasive hard-drive based DVRs.

To end-users the functionality of DVRs feels just like the rest of the "on demand" from-the-cloud experience; you can keep something for as long as the network wants you to, and then it seamlessly disappears. Collecting anything in a manner which violates the wishes of the content network is strictly verboten, and carries sentences proportional to serious criminal offences; a penalty of up to $500K and/or up to 5 years in jail per offence in the USA [10].

Privacy, Security and Obsolescence

The cloud is no longer just associated with personal general-purpose computers. It's also become a core component of many consumer goods. We now have 'smart' and 'connected' versions of almost every previously 'dumb' items: cars, TVs, watches, weighing scales, and, of course, fridges.

The benefit of having cloud connectivity for many of such devices is dubious, at best, while the potential for undesirable side-effects of such connectivity for end-users is, unfortunately, very great. These products are not only augmented by the cloud - they are now made to depend on it for even their most basic functionality. Examples of such dependence include smoke and carbon monoxide detectors that require software updates out of the box before they can function.

Unfortunately, denying such devices software updates can present unprecedented risks to end-users as well. Since these devices must remain on the network, software updates can (and often do) carry important security and reliability updates that ensure that they will remain secure from network-based attacks and continue to function [3]. Without constant updates, these devices could fall entirely under the control of remote attackers, which could compromise both end-users privacy and security. For example, attackers could gain access to video feeds of a person's home by hacking into cloud-enabled video cameras installed for security purposes. By taking control of said devices, they could also compromise the people's security; not only by commandeering control of safety-critical devices such as automobiles, but also merely by gaining privileged infor-

mation about people, such as detecting when people have gone to bed or left the house.

This dependency on 'invisible' cloud services controlled by the manufacturer can have other consequences. One is unplanned obsolescence; since online service providers disappear, making hardware depend on such services for functionality means that such devices may simply cease to operate when their corresponding supporting services go offline. Although the Internet of Things is still relatively new, (one might even say it hasn't even begun), we have not yet witnessed many such examples, yet they nonetheless remain. For example, LG recently discontinued a service required by Smart TVs they sold four years ago, leaving all those who purchased this TV suddenly without any such functionality [8].

When such termination of functionality happens suddenly with safety-critical devices, how will manufacturers be liable for putting their users' lives at risk? What will happen when Google stops the cloud services required to support their family of Nest smoke detectors, carbon monoxide alarms, and Dropcam home security cameras? Will they simply and silently stop providing their potentially life-saving functionality, or will they have the courtesy to inform their users before shutting themselves down? Trust Having appliances and electronics embedded in the most private spaces of people's lives permanently tethered to their corresponding cloud services should change the way people, as consumers, think about selecting such devices. Since these devices will necessarily have access to extremely private data about people's lives, they will need to think carefully about whether to entrust the devices and services, and the companies behind them, with such data. Will such companies respect their privacy? Will such vendors take reasonable measures to ensure that their systems will not be compromised and have their data stolen? If such companies get acquired, who will control and have access to their data post-acquisition?

Unfortunately, these additional considerations are likely to make it difficult for new start-ups to enter the commercial sector, simply because established platform providers will simply have a trust advantage. Those with an established reputation will be favoured over those without. Furthermore, with any of the dominant service providers, a person is already likely be a customer of one or more of the providers' other services, so are likely to see addition of a new device/service as an insignificant additional risk. For example, for example, if Google and a new, yet unknown startup were both offering an embedded smart device for the home, the Google would likely be significantly favoured simply because it would not introduce an additional, yet untrusted, data controller to their personal information environments.

Possible Ways Out?

As long as the "information economy" continues to be driven by personal information for its potential to be monetised in the form of targeted advertising, commercial entities offering apps, services and devices will continue to have the incentive to assume themselves as information controllers over people's sensitive, personal data. Therefore, it is extremely unlikely the loss of end-user control will simply solve itself.

However, there are several possible ways that the future could play out back towards end-user control. Core to most of these efforts are grassroot movements from the FOSS, "Redecentralize", and Maker communities.

The first is the rise of so-called "personal cloud" platforms, which constitute simple online services that people can install on and host on hardware (physical or virtual) of their

choice. The fact that such platforms are FOSS is critical to both establishing control and trust, by establishing a norm of operational transparency and a community of people to look over the code to ensure that it conforms to said norms. Transparency is also known to foster improved code quality over closed-source approaches, by inviting open critique and contribution from anyone.

Such personal cloud platforms may eventually serve a critical role in returning power to end-users. People have become accustomed to always-on, access-anywhere functionality provided by online document and data repositories, which has been shown to reduce information fragmentation, and improve people's ability to collaborate on information artefacts. Without a "cloud" of some form, however, achieving such functionality can be difficult. Therefore, introducing personal clouds under individuals' control may be a way of having said cake, while maintaining ultimate control.

Once such personal cloud platforms become widely available, the "DIY" and Maker communities may help with the hardware. These communities have already been providing kits for letting people build their own bespoke sensor and actuator hardware for assembling their own "smart devices". While such kits have not yet quite achieved the sophistication of the closed-source thermostats and security systems available off the shelf, eventually they may, especially as FOSS communities develop software ecosystems around open hardware platforms such as the Raspberry-π and Arduino.

Innovations on the software side could come as well. Since personal clouds are designed for use by a single individual, social software will need to be re-designed to work across multiple such clouds. Unlike in monolithic social platforms in which inter-personal communication essentially happens within the service, this will necessarily involve external communication between such systems. Since each will be separately managed, it is likely that such inter-communication will need to happen between different software (or different versions of the same software); as a result, to ensure continued interoperability, it will be necessary standard communication conventions will need to be established that are invariant to such differences. The W3C's Social Web Working Group [5] are perhaps foremost among organisations currently engaged in processes for defining common communication conventions to support such interactions among heterogeneous systems.

The final direction that the FOSS community might be able to help dismantle the hegemony of service providers over the world is ironically by introducing more autonomous methods of data storage in which no single entity can dominate, without destabilising the whole system. The world's first popular cryptocurrency, Bitcoin, is such an example; the Bitcoin network operates through due to the distributed contributions of thousands of computers, but due to the nature of the network, it cannot be controlled by any one of them. Beyond currencies, similar kinds of distributed, autonomous computing technologies are starting to emerge for other needs, ranging from distributed data storage (such as MaidSafe [9] and IPFS [6]) to general computation (Ethereum [7]). As these networks become more popular, people may need to rely less on centralised cloud services to support the kinds of data storage and distribution they currently turn to these intermediaries to perform.

Another possibility exists: that platform providers realise that people value choice, and seek to support its beneficial nature of to the entire ecosystem. If this happens, then these platforms may start to not only allow their offerings to interface with competitors' platforms, but to support users' in seamlessly transitioning among them. By reducing the

friction required for people to move, this would once again empower individuals to choose platforms depending on such factors as trust and quality of service, instead of being shackled to the platform(s) purely based on critical mass adoption.

There are already a handful of examples of initiatives led by platform providers towards giving people their data back. Google Takeout, for example, resulted from the efforts of the Google-internal initiative known as the *Data Liberation Front* now allows individuals to easily download all of the data they have explicitly stored into core Google services, ranging from photos stored in Google photos, to Google Drive documents, to even microdata records such as contacts, and sensed health and wellbeing data captured in the Google Fit service. We hope that the example set by Google will both be followed by other platform providers, as well as to inspire personal cloud platforms to start to support the ingestion and end-user use of such archives.

From our perspective, it is not a question of whether tools will eventually come to supporting people's' needs but when, and how; and whether such support will come from service providers themselves, or from people migrating towards more open FOSS platforms and tools that respect people's long-term needs.

References

[1] 2013. Google Shutdowns. http://www.gwern.net/Google%20shutdowns. (28 May 2013). Accessed: 2016-01-15.

[2] 2016. Google is forcing RouteBuilder to shut down. https://medium.com/hacker-daily/google-maps-is-forcing-routebuilder-to-shutdown-615ce42f413a. (9 Jan 2016). Accessed: 2016-01-15.

[3] 2016. Nest Thermostat Bug Leaves Users Cold. http://www.bbc.co.uk/news/technology-35311447. (14 Jan 2016). Accessed: 2016-01-15.

[4] 2016. Shutting down persona.org in November 2016. https://mail.mozilla.org/pipermail/persona-notices/2016/000005.html. (12 Jan 2016). Accessed: 2016-01-15.

[5] 2016. Socialwg - W3C Wiki. https://www.w3.org/Social/WG. (16 Jan 2016). Accessed: 2016-01-15.

[6] Juan Benet. 2014. IPFS-Content Addressed, Versioned, P2P File System. *arXiv preprint arXiv:1407.3561* (2014).

[7] Vitalik Buterin. 2014. Ethereum: A next-generation smart contract and decentralized application platform. *URL https://github. com/ethereum/wiki/wiki/% 5BEnglish% 5D-White-Paper* (2014).

[8] @DavidCWG. 2016. My smart TV is a few years old... so LG turned it into a dumb TV. Tweet. (12 Jan 2016). Retrieved Jan 15, 2016 from https://twitter.com/DavidCWG/status/686720545044299776.

[9] David Irvine. 2014. MaidSafe Distributed File System. (2014).

[10] Richard Owens and Rajen Akalu. 2004. Legal policy and digital rights management. *Proc. IEEE* 92, 6 (2004), 997–1003.