Managing your digital life with a Personal information management system

Serge Abiteboul, Benjamin André, Daniel Kaplan

Managing your digital life
with a Personal information management system

Serge Abiteboul
INRIA & ENS Cachan

Benjamin André
Cozy Cloud

Daniel Kaplan
Fing & MesInfos

Abstract. Computer wizards already know how to control their personal data to some extent. It is now becoming possible for everyone to do the same, and there are many advantages to doing so. Everyone should now be in a position to manage his/her *personal information management system*. This is the topic of this paper.

Context

A typical person today usually has data[1] on several devices and in a number of commercial systems that function as data traps where it is easy to check in information and difficult to remove it or sometimes to simply access it. It is also difficult, sometimes impossible, to control data access by other parties. One might consider that this is an unavoidable price to pay in order to fully take advantage of the ever-increasing amount of available information. However, this situation is not only unsatisfactory because it requires users to trade privacy against convenience but also, because it limits the value we, as individuals and as a society, can derive from the data.

We live in a world where data is considered a vital asset *and* where most people consider they have little, if any, control over their personal data. This is surely detrimental to trust, innovation, and growth.

In this world, we are also limited in leveraging all this existing information because it resides in isolated silos kept apart by technical incompatibilities, semantic fuzziness, organizational barriers, as well as privacy regulations. The situation gets worse as the number of data sources keeps growing.

Of course, users could choose to delegate all their information to a single company (some companies clearly dream of offering all the spectrum of information services). This would definitely make the users' life easier in the short run, but this would also make them totally dependent of that company. Although this is debatable, we will assume that, given a choice, most users would prefer to avoid such a solution.

Another possibility is to ask users to spend a few years of their lives studying to become computer wizards. We can safely assume that this is not what a large portion of the population craves for.

Is there another option? We believe there is one, the *Personal Information Management System* (PIMS, for short).

[1] Data that we publish (e.g., pictures), produce (e.g., contacts), coproduce socially (e.g., in social networks), data that organizations produce about us (e.g., banks, public administrations), data about us captured by sensors (e.g., gps), etc.

The personal information management system

To understand the notion of Personal Information Management System, we need to revisit today's context. Why do users “entrust” their data to services proposed by companies such as Google or Facebook? Because they enjoy using these services. Now, there are two facets to these services: they are supported by software with useful features, and they are executed on machines that are not managed by the user. What if we could separate these two facets? On one hand, a particular user would select, for each service, the best software developer or service provider that suits his or her needs. On the other hand, this user would choose a server where all these applications would run. This would therefore bring together, on a personal server, all this user’s favorite applications and all their data that are today distributed, fragmented, isolated.

This is what a PIMS does. It may look like utopia. As we will see, it is not.

The PIMS system consists of a user’s server, running the services selected by the user, storing and processing the user’s data:

- The user pays for the server (possibly owns it) so the server does what the user wants it to do and nothing else.
- The user chooses the application code to deploy on the server.
- The server software, as well as that of the services, is possibly open source (which allows for code verification on behalf of the users of the service).
- The server resides in the cloud so that it can be reached from everywhere.

Many different settings are possible. We do not need to specify a particular one. The user may own the server, or pay for a hosted server. The server may be a physical or a virtual machine. It may be physically located in the user’s home (e.g., a tv box) or not. It may run on a single machine or be distributed among several machines.

The PIMS centralizes the user’s personal information. It is a *digital home*. The PIMS is also able to exert control over information that resides in external services (e.g., Facebook), and that only gets replicated inside the PIMS. These services' business models are based on our personal data, and PIMS will not prevent them from working in this way, so long as their customers agree; however, they will need to share their data with their users, who *may* want to use them with competing platforms, or for whatever makes sense to them. PIMS do not prevent data sharing, they prevent unilateral data hoarding. The PIMS software provides the necessary support so that the user always has access to his or her information and controls (to the extent that this is possible) how information is accessed by the applications.

By centralizing access to an individual's information, the PIMS enables very useful new services that combine information from a wide variety of sources – those same silos that were prevented from collaborating together in an organizations-centric world – under the user's control and to serve his or her needs.

Is the PIMS a security risk? Of course, one could answer that it is hard to be more risky than today's large, interconnected corporate databases containing data about millions of customers, but this is hardly a comforting answer. A possible weakness is that PIMS security seems to rest on end-users when individuals have repeatedly proved to be either uninclined or unable to apply even the minimal effort towards securing their systems. However:

- The PIMS is run by a professional operator and/or on a secure hardware. Thus, security is handled by the PIMS on behalf of end-users better than it would ever be, for instance, on general-purpose personal devices such as PCs.
- The users have only to select security/privacy options under the guidance of the PIMS. The PIMS then reduces privacy risks, for example by monitoring accesses and access patterns, for all applications run within the PIMS environment.

Also, in a properly designed PIMS, each user data is strongly isolated from that of others. So, in case security has been compromised, it has been so for a single user. Pirates will therefore be more attracted to other systems with lots of data and lots of users to attack.

PIMS will not resolve the security issues for protecting users data. However, by providing a single entry point for specifying security/privacy rules, and with the support of the PIMS carefully designed with security in mind, we believe this model puts us in a better position to provide security as well as privacy to users.

Another main issue for regular users is clearly the management of their PIMS. This is where the cloud turns out to be essential. With the cloud, it is possible to have a company host the system for the users. (The user is a paying customer of that company and a contract protects the data privacy.)

PIMS are coming!

This may be observed from three different angles: society, technology, and industry.

Society is ready to move. People have had relatively little concerns so far about where their personal data goes, but this is changing for a number of reasons:

- Clear-cut abuses of massive data gathering by both governments (NSA and its European counterparts) and corporations (credit bureaus, health corporations and social networks come to mind).
- An increasing awareness by individuals of the asymmetry between what companies know about a person, and what the person actually knows about the companies (or even about herself): in Europe as well as America, consumer surveys all show that consumers are increasingly worried, not just about the *security* of their data, but also about what the organization holding data about them are likely to do with the data[2].
- A growing resentment towards intrusive marketing, cryptic personalization, and creepy " big data" inferences: As an example, according to the Pew Internet and American Life Project, 65% of US adults view personalized search as "bad" and 73% see it as a privacy invasion[3].
- An emerging understanding that personal data could be valuable to individuals as well as to corporations. "Quantified Self" applications are a case in point: millions of people seem ready to fork out $100 or more for devices that help them keep track of their health, physical shape, sleep, etc., all via data.

As a result, a series of initiatives are converging towards giving individual users not only more control over how others gather and use their personal data, but more power to

[2] As an example, see GFK,Survey on Data Privacy and Trust, 2014: http://www.gfk.com/trustsurvey/

[3] http://www.pewinternet.org/media-mentions/pew-report-65-view-personalized-search-as-bad-73-see-it-as-privacy-invasion/

actually own and use this data to their own ends. These initiatives fall into several categories:

- Privacy control: In 2009, the User Managed Access Work Group proposed specifications[4] to let an individual control the authorization of data sharing and service access made between online services on the individual's behalf, and to facilitate interoperable implementations of the specs. The current revision of privacy regulations in Europe elsewhere introduces new concepts such as "privacy by design" (data minimization, etc.), opt-in, sticky privacy policies, the "right to be forgotten", or data portability.
- Information symmetry: In the spirit of establishing a better symmetry between customers and vendors, Doc Searls and others have promoted the concept of Vendor Relationship Management[5] (VRM) since 2006. VRM emerged from the idea that customers would benefit from having an integrated view of their relationships with vendors, in the same way that vendors try to have an integrated view of their customers through CRM.
- Information ownership and use by individuals: in a 2011 report[6], the World Economic Forum wrote: "In practical terms, a person's data would be equivalent to their 'money.' It would reside in an account where it would be controlled, managed, exchanged and accounted for just like personal banking services operate today." See also, for instance, the OpenPDS Project[7] at the MIT Media Lab.

These expectations have also recently led to important personal data disclosure initiatives, such as Smart Disclosure in the US (where more than 40 million Americans can currently download and use their health data by using the same "Blue Button" on their health insurance provider's website), MiData in the UK and MesInfos[8] in France.

Technology is gearing up. Some people already use their own PIMS. They run a home server or rent a hosted server (in a 2013 market test, the French Web hosting company OVH rented 15,000 low-cost personal servers in just 10 days). They have at their disposal some rather primitive functionality, typically by developing scripts. A limiting factor is that, in order to use existing services, they have no choice but to relinquish some control over their data. For instance, if they want to partake in the social Web, they have to trust their data to Facebook or others. However, by devoting time and effort and subject to these limitations, they can manage their own data and services to some extent.

This is not for everyone, though. One needs to be highly skilled and willing to devote a lot of time in order to achieve such a result today. But things are changing rapidly:

- Abstraction technologies are helping tame the complexity of servers.
- Open source technology is increasingly available for a large range of services.
- Hardware price is now very low and the price of machine hosting has dropped.

[4] https://kantarainitiative.org/confluence/display/uma/Home

[5] Project VRM, Berkman Center for Internet and Society, Harvard University.

[6] World Economic Forum, Personal Data: The Emergence of a New Asset Class (2011), http://www.weforum.org/reports/personal-data-emergence-newasset-class

[7] de Montjoye, Y.-A., Wang S., Pentland A., On the Trusted Use of Large-Scale Personal Data. IEEE Data Engineering Bulletin, 35-4 (2012).

[8] MesInfos is a personal data disclosure experiment where several large companies (network operators, banks, retailers, insurers…) have agreed to share with a panel of customers the personal data that they hold about them.

Research in PIMS is also increasingly active[9]. A number of prototypes have been developed for storing and retrieving personal data: Lifestreams, Stuff-I've-Seen, Haystack, MyLifeBits, Connections, Seetrieve, Personal Dataspaces, or deskWeb. The tipping point appears close as illustrated by a number of projects such as Mailpile (for mail), Lima (for Dropbox-like service hosted at home), Synologie or Iomega (personal NAS), SAMI of Samsung (personal data store), and a number of self-host PIMS such as YounoHost, Amahi, ArkOS, OwnCloud or Cozy Cloud.

Large companies are getting in. PIMS also act as magnets to large companies, and in particular:

1. Traditional companies that already have large amounts of personal information. These companies, e.g. retailers, insurance companies or banks, are increasingly disintermediated from their customers by pure Internet players. They can find in PIMS an opportunity to rebuild a direct interaction with these customers.
2. Companies managing home appliances (notably Internet boxes) are natural hosts for personal information. Starting from data dedicated to specific usages, these boxes could evolve to become more generic and control more and more connected objects, services, and data.

PIMS should also be of interest to pure Internet players. Some of them, e.g., Amazon, have a great know-how in providing data services. They could seamlessly move to this new business. Others, e.g., Facebook, centered on the management of information, cannot let such a wide field of information management grow without becoming involved. However, PIMS, as we defined them here, are very far from these companies' indirect business models based on personalized advertisement. So moving in this new market would require a major change for them, and in particular, the clarification of the relationship with users (represented by the PIMS) with respect to personal data monetization.

PIMS enable new functionalities

For users, perhaps the main reason to move to PIMS is that these systems enable great new functionalities. Building of the integration of the user's data, PIMS can for instance provide:

- Global search over the person's data with a semantic layer using a personal ontology (e.g., the data organization the person likes and the person's terminology for data) that helps give meaning to the data;
- Automatic synchronization of data on different devices/systems, and global task sequencing to facilitate interoperating different devices/services;
- Exchange of information and knowledge between "friends" in a truly social way, even if these use different social network platforms, or no platform at all;
- Centralized control point for connected objects, a hub for the Web of Things; and
- Data analysis/mining over the person's information.

Conclusion

Online services have become an essential part of our daily life. However, because of them, we are all experiencing a loss of control over our personal data. With PIMS, we can regain control. PIMS also enable a wide range of new functionalities. They point towards

[9] Personal information manager, Wikipedia

a new, powerful, yet more balanced way of creating user value as well as business value. They achieve all this without giving up on ubiquity, ease of use, or security. For these reasons, we believe their benefits are so clear that PIMS will be adopted massively in a near future. What remains to be seen is what shape this evolution will take, and how it will reshuffle the cards between new "personal cloud" players, home appliance and electronics providers, established online platforms, and current personal data holders.

Will we continue to move towards an Internet dominated by oligopolies, user profiling, generalized surveillance? Will our lack of control over our data turn us more and more into passive products of a global digital economy? PIMS may be the alternative to such an outcome.