

Manage your digital life in your personal info management system

[Serge Abiteboul](#)

[INRIA](#) & [ENS Cachan](#)

Benjamin André

[Cozy Cloud](#)

Daniel Kaplan

[Fing](#) & [MesInfos](#)

Abstract

Geeks already know how to control their personal data. It is now becoming possible for everyone to do the same, and there are many advantages to doing so. Everyone should now manage his/her *personal info management system*. Such is the topic of this white paper.

Context

A typical person today usually has data 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 often to simply access it. It is also difficult, even sometimes impossible, to enforce the privacy of the data. Most countries have personal data regulations, but they are not easily enforced, in particular because data servers are often located in countries with different laws or with no regulation at all.

One could consider that this is the price to pay in order to fully take advantage of the ever-increasing amount of available information. This is not the case. The fact that the information resides in isolated silos actually complicates lives. For instance, it is not possible to perform a “global search” across multiple systems to find some information. Also, typical persons are incapable of listing the systems that host their information. Worse, they cannot even list which systems hold critical information such as credit card data. It is complicated, often impossible to synchronize the same information in different silos; and, since there is neither integration between data nor collaboration between systems, users cannot (unless they are geeks) automatize a sequence of basic operations over different systems.

The situation is getting worse because the number of data sources keeps growing. (See Box.) Individuals are facing a somewhat similar situation as the one that large companies encountered before the introduction of database and workflow systems: more and more data are available and their management becomes increasingly complex. For personal data, it got to a point where most people have lost control over their data.

Increasing dispersion of personal information

There are data that we publish (such as pictures or tweets), data that we produce (contacts, schedules, lists, etc.), data that we coproduce as the result of social interactions (on social networks, recommendation systems, etc.) or service relationships, data that organizations produce about us (as employers, schools, insurers, bankers, public administrations), some of it shared such as bills or bank statements, some of it kept away from us such as scores and other profiling data. There are data that we need to make other data systems work (such as identity or profile data). There are data that others produce about us, for example when they tag a picture. And there are more and more

data about us that software or hardware sensors capture with or without our knowledge: Web navigation, phone use, geolocation, "quantified self" measurements, contactless card readings, surveillance camera pictures, ...

This information is stored in a growing number of machines, some owned by us (phones, PCs, residential "boxes", connected devices...), some external but set up by us (a Youtube or a Facebook account), and many absolutely unknown to us. They are stored as individual files with or without metadata, as records in huge databases, coded in different ways and tagged with different identifiers.

The cloud¹ is largely responsible for the loss of control over the data that we previously could think of as "ours". There are very good reasons to adopt cloud services. Data in the cloud are ubiquitous and users are relieved from system tasks such as updating software or backing-up data. But when being relieved becomes synonymous to losing control, then the cloud becomes a problem by itself by imposing its business logic on users. In extreme cases, the external system that holds the data may simply change its usage policies, or the service may be interrupted, leaving users powerless over their data.

Can we continue to live in an information space that is growing messier and messier? Clearly, not! So what are the solutions to reach a frictionless information context?

First possibility : users can delegate all their information to a single company. It does not really matter whether it is Google, Apple or someone else. This will definitely make their life easier, e.g. they can be provided with global search, data integration and synchronization, etc. But this will also dramatically limit their freedom; so we will assume most users would prefer to avoid such a solution.

Another possibility is to ask all users to spend a few years of their lives studying to become geeks. Some of them may have the talent for this; some may even be willing to do so; but we will assume that this is not the desire of a large portion of the population.

Is there another option? We believe there is. Let's call it the *personal info management system* (Pims for short).

The personal info management system

To understand the notion of personal info management system, we need to revisit today's context.

Why do users "trust" their data to companies such as Google or Facebook? Because these companies offer fantastic services. Now, there are two facets to these services: they are supported by software with great 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, for each service, we would select the best software developer or service provider that suits our need, and on the other hand, we would choose one server where all these different "apps" will run. We would therefore bring together, on a personal server, all the users' favorite applications and all their data that are today distributed, fragmented, isolated.

¹ The expression is used to describe a variety of computing concepts that involve a large number of computers connected through a real-time communication network. In common usage, "cloud" is a metaphor for services delivered remotely through the Internet [Wikipedia].

Is this a utopia?

Material aspects of the Pims

The Pims runs “a user software” with “the user's data” on “his/her server(s)”. The important points are:

- The server is at the user’s service typically because it is paid to do that.
- The user chooses the code to deploy on the server.
- The software is open source, which is a requirement for security.
- The server is in the cloud, i.e., it can be reached from everywhere.

One sometimes uses the term *personal cloud*, to stress this last point.

We will remain voluntarily imprecise on the setting. 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 at the user’s home (e.g., a tvbox) or not. It may run on a single machine or be distributed among several machines. The software may be open source or not. The possibilities are endless, as is the range of data that Pims could store, reference and use under the user's control: Agendas, emails, contacts, medical/bank/insurance data, music, videos, books, pictures, texts, social network activity, recommendations, commercial activity, salary, taxes... *In short, all the user’s personal information.*

The Pims centralizes the user’s information. It is a “digital home”. The user selects the applications to install and the external applications to interact with. The Pims server software provides the necessary support so that the user can negotiate what data the applications require, and what they will do with the data. The centralization of all the information is an advantage: it enables great new services because it provides *individuals* with a 360° view on their lives. Furthermore, it simplifies security: It is simpler to protect a system with a single entry point than many with many entry points.

This is an essential issue that is worth detailing. There are two facets to security here: the security that is enforced by the Pims itself and that enforced by the services running on the Pims. Since we insisted that both be open source, users don't have to trust the providers of these systems. They can inspect the code (if they are competent) or trust third parties to do so. The code of the Pims is clearly the most critical. But it is only one (reasonably small) system to verify. Verifying the services is more an issue. But even if there is a security breach for one hosted service, only the privacy of the information that this service sees is compromised. In general, the information may be more secure in a Pims (a full article would be required to detail Pims’s security).

Another main issue for regular users is clearly the management of their Pims. This is where the cloud turns out to be essential. Thanks to the cloud, it is possible to have a company manage the system for the users and in particular, back up their data and make all services available all the time, from everywhere. The cloud was part of the problem; it now enables solutions.

In the spirit of Pims: VRM

Vendor Relationship Management (VRM) is the symmetric of the Customer Relation-

ship Management. Promoted by Doc Searls and others since 2006, it emerged from the idea that customers would benefit from having an integrated view of their vendors, in the same way that vendors like to have an integrated view of their customers. This would provide the users with more control over their activity and for instance, better means of comparing prices and quality of service. By establishing a better symmetry between customers and vendors, VRM is completely aligned with the objectives of Pims.

Pims are coming!

Many geeks already manage some sorts of Pims. They run a home server or rent a hosted server (in 2013, the French web hosting company OVH rented 15 000 low-cost personal servers in just 10 days). They manage to 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 relinquish some of the 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 efforts and modulo these limitations, they can manage their own data and services to a large extent.

This is not for everyone, though: today, one needs to be highly skilled and to be willing to devote a lot of time in order to achieve such a result. PCs are relatively easy to manage, however, server management is still rather cumbersome, typically based on command lines. But things are changing rapidly:

- Wonderful abstraction technologies are helping domesticate servers with virtualization and configuration management tools.
- Open source technology of the highest quality is more and more available for a large range of services.
- The price of hardware is now very low and the price of machine hosting has dropped. A hosted-low cost server could soon be as cheap as 5€/month. At that price, paying is no longer a barrier for a majority of people.

All the bricks are here. This is becoming possible. So what will trigger it?

Research in this area² is increasingly active. New results have been obtained in a number of directions such as sophisticated global search. 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. Here are some projects illustrating the trend:

- Mailpile aims at challenging Gmail. Their crowdfunding campaign reached twice the expected amount of money, paying 3 full time employees for more than a year.
- Lima, another successful crowdfunded project, raised 1.2 millions \$ to develop a device on which to deploy a Dropbox-like service, but at home.
- Personal NAS (network-connected storage) are booming, from Synologie to Iomega to "MyCloud" by Western Digital.
- Samsung recently launched the SAMI project whose goal is to provide a personal data store that will seamlessly gather all the data coming from connected devices.

² Which is Serge Abiteboul's research focus.

- And a number of new projects propose to easily self host Pims: e.g., YounoHost, Amahi, ArkOS, OwnCloud or Cozy Cloud³.

For many years, people have let companies unilaterally gather more and more of their personal data. But the tide may now be turning for a number of reasons:

- Clear-cut abuses by both governments (NSA and its European counterparts) and corporations (Facebook, Google and others' constant push towards new terms of use that restrict their user's level of control and protection; massive data gathering by websites or phone operators),
- A growing awareness by individuals of the dissymmetry between what companies know about a person, and what the person actually knows about the company (or even about herself!),
- Growing resentment towards intrusive marketing, cryptic personalization and business decisions (e.g., on pricing), and automated customer service with no real channel for customers' voices,
- Creepy "big data" inferences.

More and more customers become aware that they do not have a complete view of what organizations know about them. A smaller, but also growing number understands that personal data could be valuable to *them* as well as to corporations. These expectations have recently lead to important personal data disclosure endeavours, such as Smart Disclosure in the US, MiData in the UK and MesInfos in France.

The MesInfos project (France)⁴

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 held about them. The aim is to explore, in an agile and practical manner, the value that individuals could derive from being empowered with their own data.

Society is ready to move. Technology is gearing up. Startups are active in the field. And it also turns out that large companies are also interested.

Big players are moving in

Pims act as magnets to two kinds of large companies: traditional companies that already have large amounts of personal information, and companies managing home appliances (notably Internet boxes) that are natural hosts for personal information.

A neutral field for pre-digital companies

Pre-digital companies such as hotels or banks are increasingly disintermediated from their customers by pure Internet players such as Google, Amazon, Booking.com, Mint. In Pims, they can find a great opportunity to rebuild a direct interaction with their customers. Unlike on Internet platforms, the playing field is neutral in that each player has access to the same information. Established companies can therefore fully exploit the ex-

³ Benjamin André's startup.

⁴ MesInfos is led by Fing, of which Daniel Kaplan is CEO. Cozycloud provides the Personal servers for the experiment.

pertise they have in their particular business. Also, within Pims, they can offer new services without compromising privacy, elaborating upon the trust relationship they have built over the years.

Personal data systems offer an invaluable business opportunity to help these companies evolve in the digital era.

A natural extension for home appliance companies

There is already a huge number of « boxes » deployed either at home or in datacenters: Internet access provider "boxes", NAS servers, "smart" meters provided by energy vendors, home automation systems, "digital lockers" provided by banks, etc. These boxes are personal data spaces. However, each one is dedicated to a specific usage. They could evolve to become more generic and control more and more connected objects, services, and data. They would thereby gather more and more information. Since the providers of all these "boxes" already have a direct relationship with customers, becoming Pims seems like a natural evolution for many of them.

And pure Internet players?

Some of them, e.g., Facebook, are already focused on the management of information. They cannot let such a wide field of information management grow without getting in. However, Pims, as we defined them here, are very far from these companies' indirect business models based on personal data monetization. So moving in this new market would require a major cultural change for them.

Others, e.g., Amazon or OVH, have a great know-how in providing data services. They could seamlessly move to this new business.

Now an issue is that the business of most big Internet companies is based on extracting value from their users' data and attention. Clearly, they will not welcome the arrival of Pims. But they have to follow their users. If Pims manage to become very popular, pure Internet players will have to provide services on these systems. To do that, they will have to become interoperable with other services, and they will have to accept contracts that better protect the users' data. This will thus best serve the interests of users.

Benefits of personal info management systems

Pims carry several unique benefits:

User control over their data

Pims provide users with more control on their personal data. Users can decide who has access to what, under what rules, to achieve what goals. The key difference with the current model where the information is hosted on huge platforms, is that the user is paying for information hosting. Companies hosting the users' Pims don't need to preach "Don't be evil"; they *cannot* be evil, otherwise they run out of business.

In that sense, the information in these systems would be both better protected and more useful to individuals than in other systems.

User empowerment

Once users have selected a particular service on the cloud today, that service tries to impose on them some form of dependence, even addiction, to the service. As the saying goes, the users become the product. They lose agency and autonomy. With Pims, users are in charge. They can impose their choices. They can decide to leave a service and join a competitor. Indeed, the Pims is not here to enslave users but to empower them with the full power of their data, today limited by the friction of the silos.

New functionalities

A Pims can support new functionalities that essentially come from the existence of a unique entry point to the user's data. For instance, Pims can provide the following, unique functionalities:

- Single identity/login with automatic management of login in external services.
- Global search with a semantic layer based on a personal ontology that helps give meaning to the data. Instead of having to adapt to the terminology and concepts of each external service, the users are able to use their own.
- Automatic synchronization of data on different devices and systems, automatic back-ups.
- Automatic management of access control over all applications, e.g. by managing a single list of trusted parties.
- Global task sequencing to facilitate interoperating different devices/services. These tasks may involve services on multiple providers (e.g., Twitter and Blogger, or several banks) but be provided with a single interface on the Pims.
- 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 IoT (Internet of Things). With the multiplication of Internet objects vendors, synchronization and coordination will soon become an issue that a Pims can best address because it has access to all the user's data and devices.
- "Personal big data": Analysis of a person's personal information. Data mining may allow extracting valuable knowledge such as behavioral patterns and predictions of future actions so that the user can be more pro active with regard to health, finance, energy consumption, time management, etc.
- Apps fostering : Web services are dominated by oligopolies. That lack of diversity is obvious if you compare with the abundance of apps available for any needs on your mobile devices. Turning web services into apps is the opportunity of bring-

ing the business model (and thus the diversity) of mobile apps market place to web services.

The participation to a more “neutral” web

The big platforms such as Google, Apple, Amazon, are accumulating more and more data from their users. Thanks to "network effects", the more users and data they have, the more attractive they become to new users that bring in new data. The fact that the platforms are accumulating so much data is distorting competition between businesses in a major way. Can this be fixed? Laws and regulations are of course parts of the solution. One can also expect platforms to hand back some level of control under the pressure of customers' increasing demand for privacy. However, control only will not be sufficient to build the required level of pressure. What Pims achieve, is to point towards a new, powerful yet more balanced way of creating user value as well as business value.

By bringing back the data of each user under the control of that user, Pims can clearly make a major contribution towards bringing back more fairness on the Web.

Indeed, Pims bring more freedom to users and encourage innovation from companies in a number of ways:

- Interoperability is imposed to all services and data.
- Good practices are encouraged notably because the code is open source.
- The user can easily move a service to a new provider (portability), limiting the risk of vendor lock-in.
- The competition between services takes place on neutral ground.

What will the future be?

Online services have become an essential part of our daily life. However, because of them, we are all experiencing an unacceptable loss of control over our personal data.

Pims offer a concrete solution to that. More, they enable functionalities that the current cloud cannot offer such as single sign on, global search, collaboration across services, personal big data-style analysis, IoT hub, etc. They make all these possible without trading ubiquity and ease of use. Most importantly, they allow users to regain control over their privacy.

Will we move more and more towards a cloud 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.