The Personal Container, or Your Life in Bits

Richard Mortier, Chris Greenhalgh,
Derek McAuley, Alexa Spence
University of Nottingham
Jubilee Campus
Nottingham NG7 2TU, UK
firstname.lastname@nottingham.ac.uk

Anil Madhavapeddy, Jon Crowcroft,
Steven Hand
Cambridge University Computer Laboratory
15, JJ Thomson Avenue
Cambridge CB3 0FD, UK
firstname.lastname@cl.cam.ac.uk

Abstract

Do you record your digital tracks? Do you know whether anyone else does? Do you know what your contextual footprint is? A key research challenge for our collective digital futures is to enable people to track the information trails that they create as they operate in this, network-connected world. Given this ability, people can start to control how and to whom they expose their data, and how their data is used subsequently. This could significantly impact many businesses, including advertisers, market researchers, and service providers, in at least two directions: imposing control on those who've been able to utilise personal data they've gathered in a relatively uncontrolled fashion to date; and exposing the richness of personal data to third-parties (both individuals and companies) who've refused to exploit personal data to date, due to real or perceived barriers.

The *Personal Container* is a system to help you collate, manage, understand and exploit digital data collected by and about you. Figure 1 outlines how the Personal Container fits into the ecosystem of users, data providers and data consumers. Logically a single entity, it collects data from multiple data providers and re-exposes that data to possibly many data consumers subject to the consent of the user. In doing so it fulfils many roles:

Archive. Collating digital data about you to provide a permanent record.

Coordinator. Providing a single point of coordination for routing and translating your personal data, e.g., messages, status updates.

Guardian. Controlling exposure of that data, enabling you to exploit it while protecting it from malicious usage.

Constructing such system gives rise to several technical challenges which will be developed subsequently. In short they are:

1. How to collate and manage an individual's information across their multiple devices, datatypes and identities;
2. How to present this information to them so that they can understand and control their online exposure by granting informed consent; and
3. How to enable exploitation of collections of individuals' data by third-parties.

Many Human and Innovation challenges will also arise in the successful development of a Personal Container; these are outside the scope of this abstract and will be developed elsewhere, e.g., [2]. Some examples include:

Human Factors. How to present and visualize data stored within a Personal Container, particularly in such a way that users comprehend the implications of granting access to, and releasing, data.

Psychology. How people understand, treat, and react to privacy enforcement and information release.

Innovation. How individuals and companies can be incentivized to grant access to and make use of such a rich source of personal data.

Architecture

Although logically a single entity, a Personal Container is a *federation of instances*, each maintaining (possible overlapping) data sources. Data from a given source may be archived onto user-controlled storage, whether local or remote; or access to that data may simply be mediated by a Personal Container instance. In the latter case it is likely that the relevant instance will exist in the cloud to ensure availability. In general, instances of the following types may exist:

Third-party, single-source. These instances essentially operate as 'shims,' mapping data stored remotely by a third-party into the Personal Container. Possible examples include bank and other data with a mandated minimum lifetime.

Digital Futures '10, October 11–12, 2010, Nottingham, UK

Third-party, multi-source. These instances will be operated 'in the cloud' as a service for the user by a third-party much as, e.g., web-hosting is provided today. The hosting third-party must be somewhat trusted by the user, depending on both technology developments and the data the user chooses to store remotely.

User-operated, multi-source. Finally, these instances will exist on hardware owned and managed by the user, probably within their home or even their hand. Examples of such instances will include those running on smartphones, laptops and other commodity hardware.

However hosted, instances will be federated, providing access to different data, using techniques such as encryption, distributed indexes, and synchronisation protocols.

A key feature that the Personal Container must provide is that of *third-party access* to data.[1] Although the Personal Container is useful in isolation, it becomes dramatically more useful when the data it contains can be exploited for the user's benefit.

Research Challenges

Four key research themes arise in trying to build Personal Containers as envisaged:

Federation/Scalability

Your (logically singular) Personal Container will run on platforms from mobile phones to laptops to the cloud. It is also likely that many individuals will not want to manage their own cloud computing resource: there will be a market for third parties to run Personal Containers on behalf of consumers. The Personal Container platform must thus support federation (be able to interact with other Personal Containers and other instances of a logical Personal Container) and scalability (be able to grow as the amount of information, the types of information, and the number of users grows).

Data Management and Auditing

Personal data comes in a wide variety of types: images, audio, video, messages, locations, financial data (stock prices), live streaming; and from a wide variety of sources: self, family, friends, and colleagues. Questions to be answered include: How best to manage such diverse data? How to track and present provenance of data in the Personal Container? How to manage third-parties access to individuals' data? How to support individuals having multiple Personal Container providers (not each of which need have all a given individual's data)? How to support migration of a Personal Container between providers?

Privacy Preserving Query Mechanisms

Access to personal data has to be via some mechanism that can control information flow yet remain flexible to the querier. The approach currently being pursued involves downloading code fragments to the Personal Container to compute across one or more individuals' data before returning either some redacted data, or some aggregated result. This presents immediate system challenges, as well as longer-term formal/analytic challenges. In the first instance, the system needs to have some way to verify the safety and acceptability of such code when it is presented and executed, e.g., some code signing infrastructure. This might be done simply on the basis of a manual analysis of the code by a human. Additionally, the environment in which this code executes requires logging and analysis support to enable the data owner to observe the rate of information outflow from their Personal Container, and to intervene when they become uncomfortable with this rate. Formal metrics/techniques, e.g., Differential Privacy [1], are likely to be applicable here. Particular research elements of interest here include how to restrict the query vocabulary; how to build mechanisms to perform such queries, particularly when data may be distributed rather than simply held in one place; and how to ensure that satisfactory auditing of data access and emission can be carried out by authorities and individuals.

Automating Privacy Analysis

In the longer-term this process would clearly need automating for both static (language level) and dynamic (system level) analysis of such code fragments and their behaviour to ensure that information is not emitted at sufficient rate to cause a privacy breach. Static analyses can ensure that such plugin code cannot manipulate stored data in an unacceptable or malicious way, and perhaps cannot emit too detailed data records. Dynamic analyses can ensure that a plugin cannot emit information at too high a rate, and perhaps that multiple plugin codes cannot collude to damage an individual's privacy. Automating such analyses is a longer-term research challenge within the Personal Containers vision.

[1] This is distinct from the use of third-party services and hosting to build the Personal Container itself.

References

[1] C. Dwork. Differential privacy. In *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming, part II (ICALP 2006)*, volume 4052 of *Lecture Notes in Computer Science*, pages 1–12, Venice, Italy, July 2006.

[2] C. Mulligan and R. Mortier. Open data and competitive co-creation of value. In *Proceedings of Digital Futures*, Notttingham, UK, October 2010.

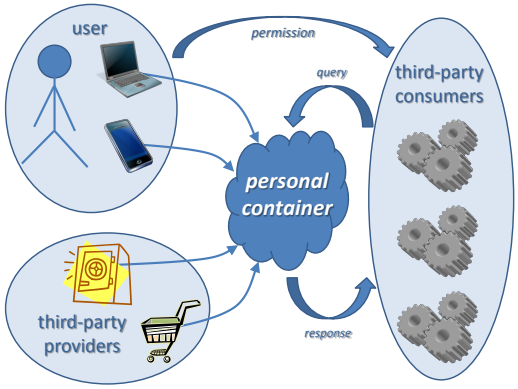

Figure 1: The Personal Container in context.

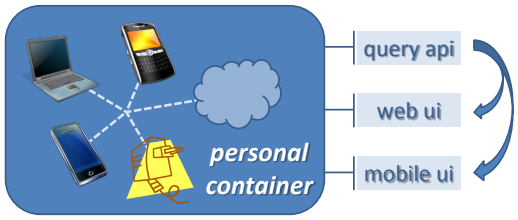

Figure 2: Architecture of a Personal Container.