# Enabling Data Subjects to Remain Data Owners

**Eliza Papadopoulou, Alex Stobart, Nick K. Taylor and M. Howard Williams**

**Abstract** Users have become used to accepting two unfortunate consequences of complying with requests to supply personal data to service providers. Firstly, the personal data that a user supplies becomes the property of the service provider, which means that the data subject loses control over what is subsequently done with their data. Secondly, provision of services is made on an "all or nothing" basis, being dependent upon the user supplying all the personal data requested by a service or forgoing use of that service entirely. We present an approach to personal data management which avoids these two unnecessary disadvantages. Personal Data Stores enable individuals to retain ownership and control of their personal data, granting service providers access to specific items of that data upon request whilst remaining the owners of their data. Trusted third parties will be required to curate the data in order to ensure that it is non-repudiatable. Privacy Policy Negotiation will enable data subjects to negotiate with service providers about how much of their personal data they disclose and how detailed that data is. Different levels of service can be provided depending on what personal data a user is prepared to disclose. In this paper we describe systems and algorithms for Personal Data Stores and Privacy Policy Negotiation which have been implemented and tested separately and show how they can be combined to the benefit of data subjects.

E. Papadopoulou · N.K. Taylor (✉) · M.H. Williams
School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK
e-mail: N.K.Taylor@hw.ac.uk

E. Papadopoulou
e-mail: E.Papadopoulou@hw.ac.uk

M.H. Williams
e-mail: M.H.Williams@hw.ac.uk

A. Stobart
Mydex CIC, Blue Square House, 272 Bath Street, Glasgow G2 4JR, UK
e-mail: alex@mydex.org



G. Jezic et al. (eds.), *Agent and Multi-Agent Systems: Technologies and Applications*, Smart Innovation, Systems and Technologies 38,
DOI 10.1007/978-3-319-19728-9_20

## 1 Introduction

Users of computer-based services are being subjected to increasing pressures to disclose personal data. These pressures take many forms, from voluntary disclosures on social media to obligatory legal disclosures required by government agencies and required disclosures which users are expected to make in order to avail themselves of proprietary services. Some of the resulting disclosures can have unintended consequences which the typical user cannot be expected to foresee, such as the transfer of their data to third parties and fourth parties, etc. Furthermore, as the number of these disclosures grows and the interoperability of services improves it will become increasingly difficult for users to keep track of, and manage, their disclosures. Laudable developments, such as greater service interoperability, can thus mitigate against an individual–centric approach to service consumption and have the potential to undermine initiatives to improve citizen empowerment such as that advocated in the European Union's Digital Agenda for Europe [1].

How have we arrived at this apparently paradoxical situation, where developments such as interoperability which should be good for service consumers could potentially be perceived as bad and, in the worst case scenario, shunned by them? The traditional service model is one where the user is expected to disclose all the information requested by a service or to forgo use of that service altogether. The choice for the consumer was a simple binary one of all or nothing. If the consumer elected to trust the service and disclose all the information requested then, again historically, the likelihood of that service provider passing their data on to others was minimal and covered by data protection legislation. In a world where most services were stand-alone and the service providers that most consumers interacted with were few in number, this was not an unreasonable model and it proved to be very effective.

However, with the advent of services which made use of other services, delivered by other providers, the traditional model began to unravel. Personal data started to be passed from one service, or provider, to another without the consumer's knowledge. Privacy policies appeared to inform consumers of what might be done with their data but they are typically so long and unfathomable that the majority of consumers do not read them. The one thing we can be sure that privacy policies have achieved is legitimising the passing of personal data from one service, or provider, to another by ostensibly obtaining the service consumer's consent. The myriad uses to which personal data could be put and the inferences that could be drawn from mining it, not least in targeted advertising, made it inevitable that a market in personal data would evolve.

Now it has been recognised that personal data has a real tangible economic value its ownership has come under scrutiny and there is a growing acceptance that the rightful owner of personal data is the subject of that data.

Yahoo's Marissa Mayer said recently that the personalised Internet "is a better Internet," emphasising: "We don't sell your personal data … We don't transfer your personal data to third parties." [1] She said users own their data and need to have

control, adding that people give up data to the government for tax assessment, social services and other purposes.

The 2011 World Economic Forum (WEF) report on Personal Data [2] stated that personal data is an economic asset class. It needs to be balanced between the needs and demands of the individual, government and private enterprise. The WEF Global IT Report 2012, "Living in a Hyperconnected World" [3], makes numerous references to the risks inherent in the flow of personal data to individual rights, privacy and cybercrime. WEF, W3C and many others have identified that a personal data ecosystem is emerging in which personal data is becoming a tangible economic asset which rightfully belongs to the subject of that data. The challenge is to devise real workable, convenient, trusted and secure systems that embody these principles.

We argue that a key pillar of individual-centricity and citizen empowerment is returning to users control over their personal data. The tools and infrastructure necessary to achieve this are now available and we present a combination of two such methods which which can readily be deployed across the Internet without the need for any architectural changes to it. In Sect. 2 we describe the first of these tools, the Personal Data Store of Mydex CIC and in Sect. 3 we present a methodology developed at Heriot-Watt University for privacy policy negotiation. In Sect. 4 we discuss some issues relating to data validation and ensuring non-repudiation of data that arise when personal data remains under the control of the consumer. We conclude and present a roadmap for the integration of the two tools in Sect. 5.

## 2 Personal Data Stores

Mydex Community Interest Company (CIC) works with individuals and organisations to enable control over how personal data is used and shared [5]. This is achieved via a Personal Data Store and a set of tools to manage identity and the consent process for data sharing online as depicted in Fig. 1 [6].

A Personal Data Store is a repository for an individual's data, not unlike the Data Box described in [7]. It is under the complete control of the individual whose date it is and is disclosed under the sole authorisation of that individual. Specific data items can be selectively released to different services. Authorisation can also be given to service providers to share particular data items with other service providers.

Mydex CIC was one of the companies originally selected by the UK Department for Work and Pensions (DWP) for ID Assurance, to provide the service now developed and operated through GDS and known as GOV.UK Verify. DWP recently consulted about new regulations for data sharing in the context of Universal Credit. The consultation seeks to find innovative ways to share personal information between DWP and local support providers, such as local authorities and housing associations. The aim is to enable these organisations to provide the

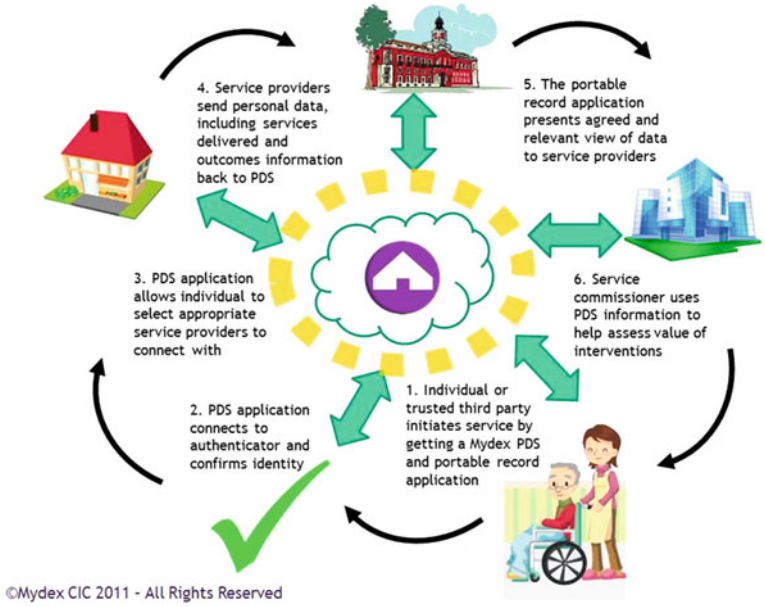


**Fig. 1** The Mydex CIC Personal Data Store and its consent process

best support for individuals claiming Universal Credit, the UK Government's new single monthly payment system for those receiving benefits and credits.

DWP is in an influential position to empower UK citizens, and to benefit hugely itself, from starting to implement a more person-centric approach to data sharing. Putting individuals in control of the data sharing process, instead of always having organisations share ever more data about them with each other, offers significant cost, efficiency and regulatory advantages. It will also help individuals acquire and prove trust, learn about, understand, and take a more active part in the digital economy.

There is a further dimension of relevance to DWP's data sharing policy. As well as providing a more elegant solution to the data-sharing challenge, connecting to Personal Data Stores means that DWP clients, notably including those with only a small amount of data about them or "thin file" clients, can then secure the required level of assurance under GOV.UK Verify for access to a wide range of digital public services.

This is much more effectively achieved when "thin file" clients are able to reuse verified evidence provided to them about their existing relationships with government departments and other organisations. This can be achieved with no loss of privacy, and without the need for changes in legislation. The data is delivered directly to the citizen, and placed under their secure control in a verified and protected format. It gives the client what is effectively a digital "proof bank"; of externally verified claims.

The same principles apply to data sharing in many other contexts: health; social care; tax; education and many other non-public services. This is an alternative person-centric approach to the current organisation-centric approach to sharing verified personal information with government agencies and service providers.

Verified data needs to have three distinct properties for it to be useable in any Identity Assurance activity –

- Certainty of issuing authority - verified organisation and issuing endpoint
- Clarity about the process of issuance or generation - this can be any number of things but it has to be understood by the relying parties' processes
- Certainty of accuracy – verification that data has not been modified since being issued

The Mydex CIC Trust Framework and platform make all of these possible with the use of cryptographic solutions to create a seamless chain of trust and a secure end to end process.

## 3 Privacy Negotiation and Flexible Levels of Service

The Personal Data Store offers a mechanism for service consumers to curate their own personal data. If a data subject only needs to deal with a small number of simple transactions then it is quite feasible for them to remain in direct control of their disclosures, perhaps manually selecting the items they wish to disclose on each and every service usage. However, even with simple transactions, there remains a need for the consumer to attach conditions to their personal data disclosures, such as the duration for which the data might be held by a service provider or to whom that provider may pass the data on. Furthermore, if the transactions require disclosure of significant amounts of data or if the data subject indulges in a large number of transactions with different services then this manual approach to approval is unlikely to suffice, requiring the data subject to approve an unmanageably large number of disclosures or to approve disclosures at each stage in a service composition or other form of interoperation. In order to assist the data subject with these issues we present the dual notions of a privacy policy and a privacy policy negotiation process [8].

A privacy policy is a document that describes how a service collects, stores, uses and disseminates the personal data of its users. More specifically, it states (a) what data are requested; (b) the purpose for which this data is requested; (c) what type of processing will be applied to this data; (d) with whom this data will be shared and (e) for how long this data will be retained. It may also include other statements about the rights of the user as well as itself with regard to the data. With the use of privacy policies, companies and more specifically services, inform their users about what happens to the users' personal data after disclosure.

Privacy Policy Negotiation is the process by which a user negotiates the terms and conditions of the privacy policy with a service. It is a solution to the "take it or

leave it" approach that is currently being implemented by millions of services worldwide. Negotiating for privacy means freedom for the user to adjust their privacy as they wish, rather than fitting the preferences of the service provider.

The privacy policy negotiation protocol defines two negotiating entities, a negotiation agent which performs the negotiation on behalf of the service provider and a negotiation client that operates on behalf of the user that wishes to use the service.

As shown in Fig. 2, the negotiation client is tasked to begin the privacy policy negotiation with the service by retrieving anonymously the privacy policy - also termed Request Policy. The service provider defines the terms and conditions for each data item that will be requested from the user, allowing different options to be defined for different data. Each data item or resource (name, age, email, contact book, location etc.) specified in the privacy policy document is defined in a Request Item element in that document. Each Request Item element defines (a) the data item it refers to, (b) the types of actions it will perform on the data item e.g. create, read, update and delete, (c) the conditions for collecting, storing and distributing the data and d) if the data item is optional.

The Privacy Policy Negotiation process involves four steps –

(a) The Negotiation Client retrieves the Request Policy document from the service provider. The Request Policy of the service provider expresses the optimal set of requirements that fit the service provider.
(b) The user is asked to configure the Request Policy with their preferred terms and conditions. A Response Policy document is generated that contains the responses of the user and is sent to the Negotiation Agent for processing.
(c) The Negotiation Agent uses a predefined set of options that indicate what options it can satisfy. It cross matches the original Request Policy with the user's Response Policy and creates a final Response Policy that includes the user's requests that it is able to satisfy and suggests alternatives for the ones that it cannot satisfy. The final Response Policy is returned to the user.
(d) If the final Response Policy contains changes made by the Negotiation Agent, the Negotiation Client presents the changes to the user and allows them to abort the negotiation or continue by accepting the service provider's alternative suggestions. If the user chooses to accept the terms, the Negotiation Client inserts the user's identity in the Response Policy, signs it digitally and returns it to the service provider. If the user chooses to abort, the negotiation halts.

When a privacy negotiation succeeds, the Negotiation Client instructs that access control rules be created reflecting the terms and conditions agreed during the privacy negotiation. The user can make use of the service at this point and the service itself will request the actual data from the user using the Personal Data Store API.

In an environment in which multiple services are combined to offer a composed service, the user would have to negotiate with all of the services separately which would be confusing to the user and would very likely drive users to avoid using such services because of the additional hassle. To avoid this, the component responsible for selecting which services to include in the service composition uses


**Fig. 2** Example of privacy policy negotiation with Edinburgh Council for using a "parking service". The privacy policy of Edinburgh Council indicates that it needs the name, age, GPS Location and activity of the user

the privacy policies and the predefined set of options of each service as another criterion for service selection. It collects the privacy policies and the predefined set of options from the available services and compares all privacy statements. For each privacy statement, it must find the option that can be satisfied by all services. This produces (a) a single privacy policy that can be used as a starting point to the negotiation with the user and (b) a modified set of options that can be satisfied by the services that will be needed during the privacy policy negotiation with the user. The Negotiation Agent of one of the services is elected to act as a delegate between the Negotiation Agents of the services included in the service composition and the Negotiation Client running on behalf of the user. This means that the delegated Negotiation Agent is authorised to perform a negotiation process without input from the Negotiation Agents of the other services. During the negotiation process, the elected Negotiation Agent has at its disposal the modified set of options that it can use to negotiate with the Negotiation Client. In a successful negotiation, the

delegated Negotiation Agent must inform the services in the composition of the result of the negotiation sending them the agreed Response Policy which they will have to adhere to when they receive the user's data.

The option for a service consumer to decide how much information they wish to disclose opens up the possibility of services providing flexible levels of service. For example, a service offering recommendations for nearby restaurants will be able to provide more tailored recommendations if the service consumer is prepared to disclose their location more precisely; disclosing location at the level of a city would result in some recommendations being a long way away from the user but disclosing location at the level of a particular street could provide a more localised set of recommendations. The notion of varying service levels to flexibly accommodate what the user is prepared to provide as input will have a very disruptive effect on the traditional service model which is based on a binary consumer choice of either telling the service everything it requests or not using the service at all.

Furthermore, Privacy Policy Negotiation, when combined with explicitly stated or automatically deduced privacy policy preferences can also resolve the age-old conflict between personalisation of services (which increases as more personal data is divulged) and privacy (which decreases as more personal data is divulged). The privacy preferences of data subjects enable privacy itself to be personalised [9].

## 4 Provenance, Validation and Repudiation

Data provenance is recognised as an important factor in establishing trust in services [10]. The validity and reliability of all data sources that contribute to decision-making processes undertaken by services has been the subject of the PROV Working Group of W3C [11]. When a data subject is given the primary role in maintaining their own data further issues of provenance arise. Trusted third parties will be required to provide assurance regarding the provenance of the data subject's data. Two key aspects of data provenance which need to be addressed are validating the data and ensuring that it is non-repudiatable.

Data validation might require access to national databases to confirm such things as date of birth, passport and driving licence numbers, etc. It is critical that such data is stamped in some way as officially verified and that the data subject is not able to corrupt that stamp. Services wishing to make use of this data can then be assured that it has not been tampered with.

Data can also change over time and it is important that data provided at one point in time, and which might have formed part of a contract on a given date, is not overwritten by an update which might change the nature of such a contract had it been entered into at a later date. For this reason data items will need to be incontrovertibly time-stamped so that, for instance, somebody getting married cannot repudiate the single status they had at an earlier point in time. Assured versioning will be essential in order to maintain the provenance of data that can change over time.

## 5 Conclusion

Personal data has now acquired real economic value and it is important that the ownership of that data is returned to the individuals who are the subject of that data. We have both the infrastructure and the tools to achieve this. We have presented one such solution; the Personal Data Store with Privacy Policy Negotiation can be provided via trusted third parties as cloud services and so be available to data subjects at any time.

The personalisation *of* privacy based on privacy preferences that we propose addresses the age-old dichotomy of personalisation *versus* privacy, by enabling data subjects to selectively disclose their personal information to different service providers depending on their context and how much they trust those providers.

In addition, our solution creates the possibility of flexibly varying service levels depending on the information which the data subject is prepared to disclose and this will move service provision on from the traditional binary choice offered to service consumers between providing all the information requested or not using a service at all.

Future work between Mydex CIC and Heriot-Watt University will develop an integrated solution based on these two concepts. Data provenance, validation and non-repudiation remain challenges but the tools to address them exist and we expect further collaborative work to identify efficacious and efficient solutions to these issues.

## References

1. de Cockborne, J.-E.: Report on an individual-centric digital agenda for Europe, JEC 130222 (2013)
2. Szalai, G.: Google chairman Eric Schmidt: "The Internet Will Disappear", The holywood reporter. http://www.hollywoodreporter.com/news/google-chairman-eric-schmidt-internet-765989 (2015)
3. World Economic Forum: Personal data: the emergence of a new asset class. World Economic Forum in collaboration with Bain & Company, Inc. (2011)
4. Dutta, S., Bilbao-Osorio, B.: The global information technology report 2012: living in a hyperconnected world. World Economic Forum and INSEAD (2012)
5. Hill, T., Alexander, D.: The third sector and the future of stakeholder engagement in challenging times - a Mydex white paper. Mydex Data Services CIC (2012)
6. Stobart, A.: Scottish government briefing – personal data stores as "Enablers of Reform". Mydex Data Services CIC (2011)
7. Haddadi, H., Howard, H., Chaudhry, A.,Crowcroft, J., Madhavapeddy, A., Mortier, R.: Personal data: thinking inside the box. http://de.arxiv.org/pdf/1501.04737 (2015)
8. Papadopoulou, E., McBurney, S.M., Taylor, N.K., Williams, M.H., Abu Shaaban, Y.: User preferences to support privacy policy handling in pervasive/ubiquitous systems. Int. J. Adv. Secur. **2**(1), 62–71 (2009)

9. Taylor, N.K., Papadopoulou, E., Gallacher, S.M., Williams, M.H.: Is there really a conflict between privacy and personalisation?, Keynote address at ISD 2011. In: Pooley R.J., Coady J., Linger H., Barry C., Lang M., Schneider, C. (eds.) Proceedings of the 20th International Conference on Information Systems Development: ISD 2011, Edinburgh UK. Information Systems Development : Reflections, Challenges and New Directions, pp. 1–9. Springer, Heidelberg, Germany. ISBN 978-1-4614-4951-5 (2013)
10. Townend, P., Webster, D., Venters, C.C., Dimitrova, V., Djemame, K., Lau, L., Xu, J., Fores, S., Viduto, V., Dibsdale, C., Taylor, N., Austin, J., McAvoy, J., Hobson, S.: Personalised provenance reasoning models and risk assessment in business systems: a case study. In: 7th IEEE International Symposium on Service-Oriented System Engineering (2013)
11. W3C: PROV-overview - W3C working group note 30. http://www.w3.org/TR/prov-overview (2013)