

Personal information data and identity: owning your personal information and it being useful.

January 2008

Michael Campbell

Graduation 2008

Interaction Design

Ravensbourne college of design and communication

Table of contents

- Acknowledgements
- 1. Introduction
- 2. Origins of personal information
- 3. Introducing the computer: the current state of personal information
 - 3a. The Internet
 - 3b. Identity
- 4. Problems with current state of personal information
 - 4a. Identity and authorisation
 - 4b. Personal Information
 - 4c. Privacy
 - 4d. Scalability
- 5. Proposed amendments to the current state
 - 5a. Identity
 - 5a (i). Social Networks/Large Provider of Single Sign-on (organisation-centric) via API
 - 5a (ii). OpenID (user-centric)
 - 5b. Authorisation
 - 5c. Communication Data
 - 5d. Personal information
 - 5d (i). Attention data
 - 5d (ii). Social network data
 - 5d (iii). Location data
 - 5e. Personal information brokerage
- 6. Conclusion
 - 6a. Individuals should own their data.
 - 6b. User-centric identification
 - 6d. Personal information brokers
 - 6e. Protected data
 - 6f. Conclusion
- Glossary
- Bibliography

Acknowledgements

I'd like to thank the staff at King's Collage Hospital for their care and hard work.

1. Introduction

Information about ourselves is becoming increasingly important. Until recently, such information has only been stored and collected on paper or mechanical systems. Now that this data is collected and stored in a digital and computerised manner the possibilities for its use have changed.

Information about us can be collected quickly, cheaply and automatically. It can then be processed by a computer programme. This can lead to improvements in our day to day lives such as ease of travel and communication. It can also lead to unfortunate consequences such as data being lost, stolen or used inappropriately. computerised systems make it possible for this to happen at a much larger scale than possible with paper based systems.

Therefore it is important that we understand exactly how our data is collected, stored and used. It is essentially that we build systems to do this which are designed, rather than simply created without consideration, as a result of our fast paced technically advancing world.

This paper seeks to examine personal information and its relationship with the notion of identity: how has it come to exist, how it is used and handled today, and where current developments will lead in the future.

It will do this by examining literature on the subject and analysing the products and services that make use of personal data and identity.

This subject has been chosen as the author considers the handling of personal information data and identity is going to be one of the biggest changes to our day to day lives that has happened since the advent of the internet.

There is the need for designers to fully understand the nature of this subject and to be involved in the designing of the products and services that make use, handle or contribute to personal information and identity.

2. Origins of personal information data

Records about births, deaths and marriages can be traced back in time to the beginning of civilisation. Records of personal finances, criminal records and trade can be traced back nearly as far.

For most of the history of civilisation information about people has been recorded and stored in manual, paper or mechanical systems. Only recently have we been recording and storing data using computerised methods.

The main holders of personal data have typically been organisations such as banks or government departments such as education or tax. In these cases it was necessary to store information about individuals for the system (bank, tax collection etc) to function correctly.

Any organisation or individual who holds data about people is referred to in the Data Protection Act 1998 (c. 29) as a 'data controller', the individual who's data is being held, is referred to as the 'data subject' and anyone who handles personal data on the behalf of the controller is referred to as a 'data processor'.

A data controller will have a filing system where it stores files that relate to individuals. The filing system could be organised alphabetically by individual's surnames or by an identifying number specific to that data subject. The organisation or data controller would use this identifying information to store information about each individual or data subject and to organise it's filing system.

An individual may wish to interact with an organisation that holds information about them. This interaction may cause the organisation to need to change the information that it stores about that person. For example if an individual wished to withdraw money from a bank the bank would need to check that they had money deposited with the bank and then change their balance on their file.

To interact with the data controller an individual would need to identify themselves. This would be done by them stating their name or identification information. The organisation would then need to authorise that this individual is who they say they are. This could be done with a signature, the statement of a shared secret or by facial recognition (photograph or familiarity between the data subject and data controller). Once the individual has been identified on the data controller's system and authenticated the individual is able to take action that changes their data on the data controller's file.

Data controllers have typically had the ability to view and in some cases edit the records they

control. They are, for all practical purposes, in charge of the data they control even though it is about someone else. The data controller would typically own the paper that the data was printed on and so if the records were stolen it would be a theft of the data controller's property.

The Data Protection Act (1998) was put in place to protect individuals from inappropriate use of their data. It recognises that there are three entities involved in personal data the data subject, data processor and data controller. The recognition of a data processor separately from the data controller is a recognition that a data controller may be so large that they trust other bodies to process the personal data of a data subject. In the case of a bank, the bank would be the data controller, the clerk employed by the bank may be the data processor and the customer of the bank would be the data subject. (Information Commissioner's Office website, 2008)

The information held on a subject's file would typically consist of identification information, security/authorisation information, and the actual personal information about the subject that the controller is intending to keep. In the case of a bank this may be information about how much money the subject has deposited with them, which would be protected data. Protected data is data about an individual that it would not be appropriate for the individual to be able to change. Other types of personal information is not protected and it is appropriate for an individual to change it if they wish.

Other information in/on the subject's file may (and usually does) include communication information (how to get in contact with subject), and operational information such as when the file was last modified or notes left by a data processor to aid in the future processing of the subject's data. Communication data should be editable by the data subject, some operational data may not be.

In some cases a data controller may exist for the sake of keeping records about individuals in a society for the benefit of society itself. For example, records about breaches of the law an individual has made (i.e. criminal record) are kept and protected from alteration by the individual or anyone who does not have the appropriate credentials.

3. Current state of personal information data

When many of the systems that stored and handled our personal information changed from being paper based to being computerised they did so with very little change to the overall model of personal information that previously existed. The systems that stored and handled personal information worked much the same way, except that they were faster and had additional functionality, such as the ability to network to other systems or databases of personal information and the ability to create back-up copies of the database quickly. The advantages of computerised

databases over paper ones are well understood. We can be sure that where personal information is held it is now at least partly (if not, totally) in electronic, computer based systems.

The prevalence of computerised systems and our use of the internet coupled with the speed, efficiency and automation of the computerised processing and recording of data has meant that the amount of data that is recorded about us has risen significantly. The advent of computerised systems for storing data had a larger impact than an increase in the speed and efficiency of storing data. It has also created opportunities for handling and processing data that did not exist before.

Companies could now look at their logistical data (data about shipping of products and stock quantities) to form conclusions that lead to efficiency gains. A computerised system could choose where to ship product X and product Y to, faster and cheaper than a paper based one. It could do this due to the amount of data it has available and the speed in which it could process it.

Retailers could use a programme to analyse data about what products were bought, where from, and when, to improve their stock buying decisions. This is an effective way of cutting costs and increasing company profits. This improvement would be made simply by analysing logistical data. This could be done simply by analysing data that already existed. What could you do if you had more data to record?

Store loyalty cards reward customers for shopping with the company associated with that card by giving customers discounts on products. The card works by associating products purchased with an individual who purchased them.

The process of signing up for a loyalty card usually involves the customer (data subject) filling in a form that will ask for a name and contact details of the customer (data subject). It may ask for other personal information such as how many people live in the individual's household, their income and other personal information that may be useful to the company (data controller) in marketing to (and potentially, profiling) their customers (data subjects).

The financial loss incurred by offering customers discounts is recovered and probably exceeded by the savings made in logistical efficiency as well as the increased profits as a result of targeted advertising, informed strategy and customer loyalty.

3a. Value in personal information data

So we introduce a slight change to the model of personal information where there are now data controllers that exist solely to create, process and handle personal information. These data controllers exist not to service any particular need or service for the data subject but as a business

unto itself. These data controllers exist because personal information has potential financial value. If personal data has financial value is there a market in it?

The Data Protection Act 1998 (c. 29) forbids personal information about a data subject to be sold or transferred to another data controller with out the permission of the data subject. Many methods that collect personal information have a small section where one can agree or disagree to having their information passed on to other data controllers. It is possible to suspect that there is a lively trade in personal information.

Trade of data (personal information or other wise) can happen either as a trade of static data (the database as it existed at a particular time) or as a dynamic of ongoing trade (access to the database over time).

It is also possible to trade personal information legally in the UK if the data that is traded is 'anonymised personal information data' meaning that the personal information can not be traced back to its data subject. This is how epidemiological research is conducted, data about a subjects health is separated from their identity and then used. It is possible that data controllers in retail trade anonymised personal information as well as non-anonymised personal information.

It is thought that most 'junk mail' is the result of data controllers passing on personal information to other data controllers. It would also be possible for company X to ask data controller Y to send all of it's data subjects mail. This could be done with out any trade of personal information. The problems of junk mail and concerns of privacy advocates has meant that companies are cautious about how they use or trade their customers personal information.

This fact may explain the apparent success of the Nectar Card scheme in the UK managed by 'The Loyalty Management Group'. The Nectar Card is another customer reward/loyalty card where customers earn points from their purchases, unlike many other loyalty cards the Nectar Card is not associated with a single store, chain or supermarket. It is a reward card that is used by data subjects across a wide range of retail outlets and services. An individual can use their Nectar Card in restaurants, petrol stations, department stores, opticians, on breakdown cover, home utilities such as gas, electricity and communications, car hire and travel anywhere where the provider is part of the Nectar Card scheme. Nectar Card holders can also chose to exchange more of their personal data for points by filling in surveys on the Nectar website, points can then be used to pay for services.

With a large database of consumer habits, incomes, wants and needs across a wide range of goods and services it could be possible to profile customers and to market to them more effectively than you would with a record of purchases from a single retail outlet.

3b. The internet

The amount of data that is recorded and exists about people today is considerably higher than it was before the wide spread use of the internet. We use the internet to store and organise our personal information such as our relationships to other people, activities, and things we find interesting. We also use the internet to publish a great deal of personal information such as in blogs (web logs) or via social bookmarking tools. More and more of our work and school related activities are also recorded in live on-line working documents (e.g. Google docs).

We do this with a wide range of services that each hold a some of our personal data. We register with and then sign-in to many services which have much the same model of personal information data and identity discussed earlier. To use a web service it is usually required that individuals register themselves with that service. To register with a web service (data controller) the user (data subject) will normally have to volunteer at least some information about themselves, normally just an email address (communication data), the subject may also be asked to choose a unique (to the data controller) 'user name' which may double as that subject's identity, and a password as a shared secret type of authentication. When the user wishes to use the web service the user tells the web service who they are (user name), the system checks to see if it has a record of the user name, the subject is asked for the shared secret, the system checks that it corresponds to the authentication data in the subject's file and if it does, the subject can use the service. Much like discussed earlier with paper based systems, except in this case, the data processor is a computer programme not a human.

In the early days of the web the main reason an individual would want/need to register and create an account with a website or service was to enable them to make a purchase. Now in a so called 'web 2.0' (O'Reilly 2007) world the reasons we would want/need to log into a web service have significantly expanded.

Web services exist today that do a wide range of things; uploading/watching video, searching, bookmarking, calendaring, micro-blogging, social news ranking, uploading/organising/viewing photos, the list goes on. It can be said that the function of these web 2.0 services and applications is to allow users (data subjects) to create content or to create data about content (meta data, see glossary) that is viewable online.

The way that web2.0 has grown has meant that there are many services that each specialise in their own particular type of content or content meta data. weather it be blogs, photos, videos, social networks, messages or bookmarks. This means that it is possible for an individual (data subject) to have their content and meta data (personal information data) scattered across many different

services (data controllers). It is also possible that that individual may need to have different identities for each service they use. When signing up to each service it is likely that the individual will have to enter the same information, such as communication data and (as is the nature of web2.0) data about who they are friends/contacts with/of. If an individual uses more than a few of these services they may find themselves doing this many times. They will also need to alter this data in many places if they wish to keep it current.

3c. Attention

So far we have talked about the advantages for retailers to analysing personal information data for the purpose of increasing financial profits: e.g. how much expendable income people have and what they spend it on. This can enable retailers to market more effectively to customers, because they can market to them only what they are statistically likely to buy.

With the increasing in use of the internet there has been an increase in the amount of operational data that relates to individuals. Operational data on a web service's records may include what links have been clicked on as well as which user clicked on them, and when. Together it is possible for a web service to conclude how long a particular web page was viewed. This operation data can indicate how a web user might be allocating their attention - a valuable commodity.

The attention economy has existed as long as, and played a part in, television, radio and print. This economy is based on the quantity of expendable attention people have and how they spend it. However the importance of this economy has increased and is playing a much larger role in recent times. This is due to both the business model of many of the web services that we trust with our data, as well as the ability to record indications of attention on the internet medium. (Davenport and Beck, 2001)

Most popular web services (data controllers) are free for the user (data subject) to use. The way that they generate revenue is by charging third parties (advertisers) to show adverts to the users of their site. Data about where an individual has spent attention and what they find worthy of attention can be very effective in targeting advertising effectively.

A good example of a clever use of this attention data is the search engine Google. Google's business model is one that demonstrates the relationship between attention and high return on investment for advertisers. Before Google was a business that made money it was just a search engine. The user gave Google a search query and Google would look at its database of the internet and show the user a list of web pages that were relevant to that query. Pages were ranked according to their relevance to the search query and the PageRank algorithm (Page et al, 1998). Unlike many other websites at

the time Google didn't have any advertising on their web page. The adverts that other websites had on their page operated in much the same way as the advertising model for print. What a company would pay for an advert in a magazine was based on the expendable income of its readers and how many copies of the magazine were sold.

The advertising model that Google now operates charges advertisers based on how much they are willing to pay and only if a user clicks on their advert. An individual will only be shown an advert on Google once they have made a search query. Adverts are then selected from a database based on their relevance to the search query. The adverts are considerably cheaper (to the advertiser) than they would be in the previously mentioned advertising model but they typically have a higher return on investment. This may be because an individual is more likely to buy a product from an advertiser if they saw an advert for that product while searching for it.

Space is not the same premium it is in the print advertising world. Google can charge less for each advert but can show a potentially infinite amount of adverts along the 'long tail' of the graph. (Anderson, 2006). There is possibly an infinite amount of different search queries that an individual can enter into a search engine and therefore there is the potential for there to be an infinite amount of space for the search engine to show the individual an advert.

The adverts that an individual sees is related to how the user is intending to spend their attention. The user isn't having to ignore adverts that are trying to distract them from what they were doing, the adverts are offered to the user by Google as something that could be what they are looking for.

Google can use personal information data which is given to Google (the user's search query) to power their advertising model. Operational data about what link was clicked on as a result of what search query can be used to improve their search algorithm. (Battelle, 2006)

It could therefore be argued that users pay to use Google with their personal information.

3d. Identity

User names are more than just a method of identifying yourself to a single organisation (data controller) they are also used by individuals to identify themselves to each other. Users of social networks or web forums may refer to each other by their user name rather than their real name. In some cases these user names span more than one web service and individuals use the same user name on web service X as they do on Y. This way other users of both systems are led to believe that it is the same individual on both systems, although there is no way of knowing for sure. It may also be that individuals have several user names that they identify with.

4. Problems with current state of personal information

The current model of personal information and identity is an evolution of a model that existed using a fundamentally different medium (paper). The processes and methods that we have inherited from this model may not be appropriate in this new age where personal information and identity are used/handled in a drastically different way and through the medium of the internet.

4a. Identity and authorisation

One common complaint from users (data subjects) about their current interactions with web services (data controllers) is that they have to remember too many different user names (identifying data) and passwords (authorisation data).

If an individual is concerned about the safety of their accounts with the web services they use, they may choose to use different user name and password pairs with each web service. Meaning that a user may have difficulty remembering what their login details are for a particular service. There is software that exists on personal computers that can store an individual's log-in details for various websites, but this is not a solution to the problem. An individual may need to log-in to a service while on a different computer or there may be more than one individual that uses a single computer. It may also be the case that the user name (or on line identity) that an individual usually uses to identify themselves to other users of a social service has been taken by another individual.

Essentially there is no effective method for identifying oneself on the internet. The method of identification that we have inherited is organisation-centric. This was appropriate when the only time we needed to identify ourselves was to a few organisations. Now, however we want/need to identify ourselves to each other and to many different organisation systems. The current model of personal information and identity means that information about us is scattered about the web in 'dislocated silos' (Forrester, 2007).

4b. Personal information

The Data Protection Act is a valuable piece of legislation, however alone it will not ensure that our personal information is kept safe. The pace at which the systems that hold our personal information data have evolved has meant that it is only possible for the Data Protection Act to make generic statements rather than specify particular methods for handling data.

Without there being specific data handling methods which should be adhered to by law it is possible for data controllers to store and handle personal information data in ways that are insecure. These insecure methods of storing personal information can lead to personal information data being used

in illicit ways (Privacy International, 2007).

The Data Protection Act allows an individual to have access to their data via a written request. However for data to be truly useful to an individual in the same way it is useful to a data controller it will need to be given to an individual in a computer readable way and faster than a written request will allow.

There is currently now way in which an individual can request and achieve the exchange of data from one service to another. An individual can not for example move their attention data from one online retailer to another.

Until individuals have on demand access to their personal data in a computer readable format they will not have as much power over their own data (which is effectively the result of a creative partnership between the individual and the web service) as the controllers that hold it.

4c. Privacy

The privacy threat on the Internet arises from a number of factors. Increasing disclosure by consumers of personal information allows companies to capture and process data to a significant extent. New technologies permit the capture of increasingly detailed levels of information. Meanwhile, new Internet products often involve a requirement for user registration, enabling of identifying techniques and agreement to terms and conditions that are frequently hostile to privacy. (Privacy International, 2007, A Race to the Bottom: Privacy Ranking of Internet Service Companies)

Some web services are making an effort to ameliorate the concerns of privacy advocates, such as allowing individual to delete or even edit the personal information that is held on their system. Google for example allows users to view, and delete their search history via Google's interface. This isn't enough for Privacy International, they believe that Google is able to keep information for too long and that it has the opportunity to know too much about its users.

4d. Scalability

As we move closer to a world where the boundaries between the internet and our day to day lives becomes increasingly blurred; towards a world of ubiquitous computing. We need to design a system of identity and personal information handling that not only solves some of today's problems but also scales well into the future.

5. Proposed amendments to the current state

Amendments and proposals have been made by various parties as an attempt to improve the way that personal information and identity are handled.

5a. Identity

There are several trends that may eventually lead to at least a partial solution to the problems with identity previously discussed.

5a (i). Social Networks/Large Provider of Single Sign-on (organisation-centric)

In a similar way that the Nectar Card scheme aims to be a single organisation where individuals' identities are aggregated across all member organisations, it could be that a single social network or large web service provider could become an aggregation of identity across all web services. Many web users have accounts with one of the few largest companies on the web, these companies can allow their users to sign in to many of their individual web services and partner sites with a single identity¹. This could be expanded so that all web services could use an identification Application programming Interface (API, see glossary) to handle identity from one of these large social networks or web service providers.

An attempt of this type has been made before with 'Microsoft Passport', passport is not seen as being successful for several reasons; users found its interface difficult, people did not like the idea of Microsoft being the controller of their identity, in addition an number of implementation failures. (Harrison, 2006, p.6).

5a (ii). OpenID (user-centric)

The OpenID project is an attempt to establish a user-centric system of identity for web services. This means that the user is the centre of the model and that each individual has their own identity separately from any organisation they subscribe to.

An OpenID provider is an organisation that handles an individual's on-line identity. Each individual has their own Universal Resource Locator (URL, see glossary) that points to their OpenID provider's server. An individual may also have other URLs that contain information about where that URL's OpenID provider is. This allows users to keep the same identity but change their OpenID provider if they wish.

With the OpenID system an individual uses their OpenID URL to log-in to a web service (referred to in OpenID terminology as a 'relying party'), the web service then asks the OpenID provider to check that the user is who they say they are. The OpenID provider then checks this by authenticating the user (see section 5b). If the OpenID provider believes the user is who they say they are, it will then send the relying party a confirmation message(openid.net, 2008).

¹ Microsoft passport/Live ID and Google and it's Google applications for example.

OpenID version 1.0 also contains a method of communicating some personal information such as communication data as well as identity and authentication. With version 2.0 it may be possible to exchange any item of personal information.

5b. Authorisation

There are a range of different methods that can be used to authenticate that an individual is who they say they are when using a web service. As discussed earlier, it can be done with a signature or by facial recognition if the individual is somewhere there somewhere there is someone to confirm or deny authorisation on behalf of the data controller.

In the case of web services where there is a layer of technology that separates you from any individual who can confirm or deny that you are who you say you are, it is necessary to use different methods. Shared secrets such as passwords have already been discussed. Key chain fobs which generate a pass number is another method as is so called 'out of medium' verification. This works by sending a password to the user via non-web based method such as mobile (short message system) SMS. There are also several types of certificate authentication including Microsoft CardSpace although specific web authorisation technologies are beyond the scope of this paper.

5c. Personal Information

To facilitate the portability and owner control of an individual's own personal data, it is necessary to standardise how that data is stored. An open standard is one that is not owned or controlled by a single entity but by a community. This means that it is not possible for a single entity to exclude other entities from using it.

Standard methods of transferring personal data also need to be created. With standard computerised ways of storing and transmitting personal information data it is possible to build computerised tools allow us to own and control our own personal data.

5e (i). Attention data

One such proposed standard way of storing personal information data is Attention Profile Mark-up Language (APML).

APML allows you to share your own personal Attention Profile in much the same way that OPML [Outline Processor Mark-up Language] allows the exchange of reading lists between News Readers. The idea is to compress all forms of Attention Data into a portable file format containing a description of your ranked interests. (APML.org, 2008)

APML is a method of storing how an entity wishes to spend their attention, based on observing the

actions and implicit instructions of its user. APML is designed to be read by computers so that programmes which are capable of measuring indicators of an individual's attention can use and contribute to their attention profile.

APML is an eXtensible Mark-up Language file (XML, see glossary) that's body contains two sections, implicit data and explicit data. Explicit data is set directly by the owner of the file and implicit data is set by computer programmes. In each section there are source items and concept items. Source items are places where information potentially worthy of attention are found, e.g. RSS Feeds. Concept items are things that are potentially worthy of attention, e.g. keywords. Each item is ranked according to how likely it is to be worth the individual's attention.

Platforms where attention is spent can also measure an individual's attention. An individual's web browser could measure how long the user spent reading a particular web site and if the user clicked on any links while there and write this information back to the individuals Attention Profile.

5d (ii). Social network data

Fried of a friend (FOAF) is an open standard which contains information about an individual's social network, i.e. who they know and what their relationship is with them. FOAF is XML based, it also makes use of the W3C Resource Description Framework (W3C, 2008) which is a semantic way of addressing resources on the internet (Berners-Lee, 2001). It works by an individual having a unique URL that only they control. This URL acts as an identifier of that individual. Within the web page found on that URL is code which links to a file that is an individual's FOAF file. A computer can read an individual's FOAF file and use this data in a social web application.

With wide scale adoption of FOAF data our connections to each other on a large scale will be for us to use and examine rather than only being available to a few large internet based companies.

The Open Social API is being developed by a group of interested parties as a method for using an individual's social network data from a social networking service and then remixing or using it within another web service. With social networking sites opening their social graph to be used by other web services it may be possible to put an end to the frustration users feel when having to enter their social data inside every web service they use. It may also go some way toward giving individuals some sort of web identity that is transferable from service to service.

5d (iii). Location data

Fire Eagle is the working title of a research project by Yahoo that aims to be a broker for

information about an individual's location past, present, and future.²

At the time of writing the project is in closed alpha testing, therefore details of how exactly it operates and what it does are not available. In a talk at the Future of Web Applications conference 2007 Tom Coates of yahoo spoke of his work on the project (Coates, 2007, FOWA).

Coates outlined that Fire Eagle receives updates about an individual's current location via a range of mostly mobile phone related technologies. It then manages requests from other services that would like to know where the individual has been, is, or will be. The individual can decide which web services are entitled to what information as well as how accurate that information is. When an individual is in their home Fire Eagle can tell web services the individual is in their home, and not the longitude and latitude coordinates of where that is.

This data could be used in a number of ways to improve the experience of the individual, it could potentially give a location based context to search results. Needless to say, data about an individual's location need to be handled very carefully. However the ethical issues and potential uses of this particular type of data are beyond the scope of this paper.

5e. Personal information brokerage

In two papers by Eidentity Ltd the issues of personal information brokerage and user-centric ID are discussed in relationship to the proposed Managing Information Across Partners (MIAP) programme (Harrison, 2007) (Harrison, 2006).

Personal information brokerage is a user-centric model of personal information data. In this model an individual commissions an information broker. That information broker then handles the personal information of the individual on their behalf. Services that the individual uses can then subscribe to their information broker where the broker will serve personal information to the service subscribing (relying party) if the individual authorises it.

...[personal information brokerage] place[s] the control of data into the hands of the individuals who created it, enabling them to collect merge and sell their data in a new marketplace. In addition to giving customers a share in the value of their own data, businesses have the opportunity to access a more complete picture of their customers, that would otherwise be in breach of the Data Protection Act or too expensive to collate. (live|work studio Ltd, 2004)

² <http://fireeagle.research.yahoo.com/>

6. Conclusion

6a. Individuals should own their data

When a musician creates a song in the UK that song is the legal intellectual property of the musician. The owner of intellectual property has legal copyright of that property. Meaning they can legally dictate how their property is copied, published, distributed and any profit made from that property is legally the creators. The same is true of intellectual property; music, literary works, art and computer code. It can be argued that this should also be the case for data that is created by individuals using or participating in a system. This applies to content and meta data that individuals publish online unless the individual specify otherwise.

If an individual (data subject) did not use a service (data controller) then the data that it records would not exist. Neither would it exist if the service did not record it. Therefore the individual and the service are co-creators of that data. Consequently the service and the individual should have legal control of that data. This paper proposes that they should have the legal right to dictate how the data is copied, published, and distributed. Any profit made from that data should be split between the creators.

With legal ownership of their data individuals would be able to dictate how their data was used. The co-creator of that data (data collector) would have the legal right to use that data for it's own end as long as it did not violate the rights of the individual.

The author believes that if an artist has copyright of their work so too should the creators of data. However, this alone will have limited impact on the problems discussed in this paper. The process of giving an individual legal ownership of their data (as well as implementing the technology that enables that) needs to be carefully considered and designed.

Furthermore, some may argue that the Data Protection Act is enough to ensure that individuals have control of their own data, others would disagree. If an individual can not get access to their data in a machine readable format then it is not as useful to the individual (data subject) as it is to the service that holds it (data controller).

6b. User-centric Identification

One possible solution to this problem would be a personal information brokerage system built on top of a standardised user-centric identification mechanism.

OpenID stands a good chance of bringing an improvement to the current model of identity and

personal information. If individual's had their own URL as a method of identifying themselves it could be possible to build systems that allow our personal data that is collected on the services individual's use to be communicated back to them.

This combined with the movement towards designing open standards for storing and processing different types of personal data (attention data in APML and social data in FOAF as well as others) means that our personal information can be communicated to us in a utilizable way rather than via the paper method that the Data Protection Act dictates, which affords relatively little utility at all.

With our personal information being held on systems that we have control over we would be able to dictate how it is used and how long it is stored for. Communication data, preferences and profiles could be stored in a single place that all web services update their records from. OpenID can also be used as a tool for individuals to identify each other when using the internet.

6d. Personal information brokers

The author believes that we could see the rise of so called personal information brokers who act on the behalf of individuals to manage their on-line identity, store thier personal information and also to process that information for them.

These information brokers would allow us to control who has access to our data and how it is used. They would process it on our behalf to spot patterns in our data from different sources and to form conclusions that we may find useful.

This collation of our personal data in one place not only goes some way to solving some of the problems mentioned previously (privacy etc) but can also facilitate new business. Information brokers could become an advertising platform.

Personal information brokers could be the foundation of a new information exchange market where the value of an individual's personal information is exchanged for, or contributes to, the purchase of goods and services to the benefit of the individual. This is made possible by the opportunities for highly targeted advertising and the potential increased return on investment this brings.

6e. Protected data

Is the trust based method of choosing an identity provider enough to protect individual's from the illicit use of their data? An alternative would be the introduction of specific legislation that builds upon the principals of the Data Protection Act to ensure information brokers are held accountable

for the responsible management of individual's data. Before this legislation takes place there needs to be consideration of whether or not there should be a web based method of identifying a subjectd that can be legally ratified. OpenID as it is currently, is not designed to be such a method.

Information brokerage has the potential to be. However it needs to be fully understood and tested before any new legislation is introduced.

Further questions we will need to address include:

Does society need to be able to associate an online identifier with a real individual person for the purpose of protected data (criminal records, Bank balances, etc)? Or should this this data continue to be held inside systems inside organisation-centric models of identification?

6f. Conclusion

In conclusion, this paper has demonstrated the value of and possibilities for user-centric identification. It has identified the current challenges to it's implementation- technical, social and legislative. It is clear that if user-centric identification is to become a reality these challenges will need to be addressed by technical designers and legislators alike.

Glossary

Personal information: is data that is about or relates to an individual or their actions.

Shared secrets: are a method of verifying that an individual is who they say they are. It works when both the data subject and the controller share a secret. A Password is a type of shared secret authentication. It can fail when someone other than the data controller or the data subject discovers the secret.

Web service: is more than just a web site, it is a service that operates behind a web site.

Meta data: is data about content. Examples include information about when the content was created and who by. It can also be data towards a so called folksonomy, keywords or tags that describe data.

OpenID provider: it the entity who provides the technology that is an OpenID, This can be a organisation that controls many OpenID, or it could even be the individual who provides their own OpenID.

Application programmeming Interface (API): are methods for allowing third party developers to use aspects of a system that an individual has created.

Universal Resource Locator (URL): is a way of addressing resources on the internet.

eXtensible Mark-up Language (XML): is a method of storing data and it's relationship to other data in an ontology.

Bibliography

Anderson, C. (2006) **The Long Tail: Why the Future of Business is Selling Less of More.** Hyperion

APML.org (2008). [online]. Available at <http://www.apml.org/> (accessed 16/01/08)

Battelle, J (2006) **The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture.** Nicholas Brealey Publishing Ltd.

Berners-Lee, T. (May 2001). 'The Semantic Web'. **Scientific American**

Coates, T, (2007), 'An Insight into FireEagle' **The Future of Web Apps Expo.**

Data Protection Act (1998). [online]. Available at http://www.opsi.gov.uk/acts/acts1998/plain/ukpga_19980029_en (accessed 16/01/08)

Davenport, H and Beck, C. (2001). **The Attention Economy.** Harvard Business School Press.

Forrester, Ian. (2007). **Pipelines plumbing for the next web.** [online]. Available at <http://www.cubicgarden.com/blojsom/blog/cubicgarden/xml/Semantic+web/2007/05/16/Plumbing-for-the-next-web-at-Xtech-2007.html> (accessed 16/01/08)

Harrison, J, (2007), **MIAP and PIB – the 'Wider Picture': A response to comments by DfES.** Newbury

Harrison, J, (2006), **House of lord Science & Technology committee Investigation into Personal Internet Security, Submission of evidence by Edentity Ltd, November 2006.**

Information Commissioner's Office Website (2008). [online] London. Available at http://www.ico.gov.uk/what_we_cover/data_protection.aspx (accessed 16/01/08)

live|work studio Ltd (2004) **Loome: Personal Data Brokering**. [online] Available at <http://www.livework.co.uk/projects/loome/> (accessed 16/01/08)

Openid.net, (2008). [online] Available at <http://www.openid.net/> (accessed 16/01/08)

O'Reilly, Tim. (2007) **What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software**. MPRA Paper 4578, University Library of Munich, Germany.

Page, L., Brin, S., Motwani, R., Winograd, T., (1998). **The pagerank citation ranking: Bringing order to the web**. [online] Stanford. Available at <http://dbpubs.stanford.edu:8090/aux/index-en.html> (accessed 16/01/08)

Privacy international (2007). [online].

'**Privacy International to pursue data breach legal action against UK government**' available at [http://www.privacyinternational.org/article.shtml?cmd\[347\]=x-347-558703](http://www.privacyinternational.org/article.shtml?cmd[347]=x-347-558703) (accessed 16/01/08)

Privacy international (2007). [online].

'**A Race to the Bottom: Privacy Ranking of Internet Service Companies**' available at [http://www.privacyinternational.org/article.shtml?cmd\[347\]=x-347-553961](http://www.privacyinternational.org/article.shtml?cmd[347]=x-347-553961) (accessed 16/01/08)

Today programme. (2007). BBC Radio 4. 11/07/07

W3C, (2008) **Resource Description Framework**. [online]. Available at <http://www.w3.org/RDF/> (accessed 16/01/08)