Submission to the W3C Workshop on Web Tracking and User Privacy 28/29 April 2011, Princeton, NJ, USA

By: The Personal Data Ecosystem Consortium

Kaliya Hamlin, Executive Director, Personal Data Ecosystem Collaborative Consortium
director@personaldataecosystem.org @identitywoman Mobile: 510-472-9069

Mary Hodder, Chairman, Personal Data Ecosystem Collaborative Consortium
mary@personaldataecosystem.org Mobile: 510-701-1975

Submitted April 1, 2011

I just learned about this workshop at a dinner yesterday and I am submitting the position paper accordingly. I believe view we represent is vital to bring into the discussion as a better approach to user privacy. I hope you will still consider our position for inclusion in the workshop.

Personal Data Storage and Services: A Middle Way between Do Not Track and Business as Usual Tracking

The Personal Data Ecosystem Consortium represents a community of end-user advocates and technology innovators focused on individual rights and access to individuals' own personal data, and the business and innovation opportunity that this new user-management and control.

The perspective that we have is quite different then either Do Not Track or Business as Usual Tracking. Personal Data Storage and Services where individuals effectively “stalk themselves” aggregating their own data streams from diverse sources into a personal data service.

The scope of activity in this space goes beyond web surfing habits and tracking data. We chose to reflect this broader scope in this position paper because we we know services are being built beyond that scope there are over 20 startups are building today for this emerging ecosystem and large firms are doing incubator projects testing the waters and the World Economic Forum just released a report articulating the value of Personal Data as an Emerging Asset Class.

We believe this approach has great promise because besides empowering users it also represents new business opportunities and incentives for companies who currently are dependent on the information they glean from a whole range of online tracking techniques to have access to even better information about the people they are seeking to market to.

The Two Ends of the Spectrum

On one end of the spectrum is the “Do not track” view, which relies on using technology and a legal mandate to prevent any data collection (as per the FTC Proposal). In this scenario, cross site behavioral targeting is suppressed because users signal they do not want any information to be collected on them as they move about the web. In this approach the economic value advertisers have been getting through higher click-through rates by providing more targeted ads is eliminated and sites that receive revenue from serving targeted ads is reduced if not eliminated. The economic value of the data is not captured by the end-user nor is it benefiting the media/advertising/data aggregating complex.

On the other end of the spectrum is the mode where we leave “Business as usual” in place as it has developed over the last few years. The door is wide open for ever more “innovative” pervasive and intrusive data collection, tracking and cross referencing for behavioral targeting in developing profiles -- digital dossiers created on billions of people, without their knowledge or consent, based on IP address, device identification, e-mail address etc. The status quo is highly invasive of people’s privacy, linking their activities across contexts they wish to keep separate or private if they chose to do so. In addition, decisions about people's lives are increasingly made from such data, and they are not aware of it, though the consequences can be quite severe. Economic value is derived, but at the expense of the basic dignity and privacy rights (ie personal control) of the individual.

Personal Data Storage and Services

Personal data storage services are emerging, representing a middle way through, to provide an opt-in modality with greater choice and control to the individual over their data AND offer greater economic value to the business community, with huge innovation and market opportunities. This market, we believe, will be much larger than the current one based upon surreptitious stalking, and be based upon an ethical model involving the user in the transactions the might occur with their data, where choice, transparency, access and control are central features for users.

As envisioned, Personal Data Storage Services (PDS) allow individuals to aggregate their personal data, to manage it and then give permissioned access to businesses and services they choose -- businesses they trust to provide better customization, transparency, access and the ability to correct, as well more relevant search results and commercial offers, resulting in increased value for the user from their data.

Over the last year, activity in this space has grown tremendously. In this emerging field of innovation, we have identified over thirteen startups (some of them with significant venture capital funding), at least three open source projects, several technical standards efforts in recognized international standards organizations along with companies in the web, mobile, entertainment and banking industries working on this model.

One of the most important things about this emerging space is that it has engendered active business development both in the United States and across Europe. In other words, this model is viable across North American and European privacy regimes. Furthermore, the PDS model offers the possibility of achieving global interoperability, one of the key goals articulated by the Commerce Department for this forthcoming set of policies and regulations.

People are the Only Ethical Integration Point for Disparate Data Sets

Today there is a personal data ecosystem emerging in which almost everyone unknowingly participates but without the personal individual controls to afford user-centric privacy. People unwittingly emit information about themselves, their activities and intentions, in various digital forms. It is collected by a wide range of institutions and businesses with which people interact directly; then it is assembled by data brokers and sold to data users (ie businesses that exploit our data without including us in the transaction). This chain of activity happens with almost no participation or awareness on the part of the data subject: the individual.

We believe that the individual is the only ethical integration point for this comprehensive and vast range of disparate personal data. For example, the list of data types below was put together by Marc Davis for the World Economic Forum talk: Re-Thinking Personal Data event in June of 2010. It highlights the vast range of datasets about an individual that might be in some digital form in some database somewhere.

Identity and Relationships:
* Identity (IDs, User Names, Email Addresses, Phone Numbers, Nicknames, Passwords, Personas)
* Demographic Data (Age, Sex, Addresses, Education, Work History, Resume)
* Interests (Declared Interests, Likes, Favorites, Tags, Preferences, Settings)
* Personal Devices (Device IDs, IP Addresses, Bluetooth IDs, SSIDs, SIMs, IMEIs, etc.)
* Relationships (Address Book Contacts, Communications Contacts, Social Network Relationships, Family Relationships and Genealogy, Group Memberships, Call Logs, Messaging Logs)

Context:
* Location (Current Location, Past Locations, Planned Future Locations)
* People (Co-present and Interacted-with People in the World and on the Web)
* Objects (Co-present and Interacted-with Real World Objects)
* Events (Calendar Data, Event Data from Web Services)

Activity:
* Browser Activity (Clicks, Keystrokes, Sites Visited, Queries, Bookmarks)
* Client Applications and OS Activity (Clicks, Keystrokes, Applications, OS Functions)

* Real World Activity (Eating, Drinking, Driving, Shopping, Sleeping, etc.)

Communications:
* Text (SMS, IM, Email, Attachments, Direct Messages, Status Text, Shared Bookmarks, Shared Links Comments, Blog Posts, Documents)
* Speech (Voice Calls, Voice Mail)
* Social Media (Photos, Videos, Streamed Video, Podcasts, Produced Music, Software)
* Presence (Communication Availability and Channels)

Content:
* Private Documents (Word Processing Documents, Spreadsheets, Project Plans, Presentations, etc.)
* Consumed Media (Books, Photos, Videos, Music, Podcasts, Audiobooks, Games, Software)
* Financial Data (Income, Expenses, Transactions, Accounts, Assets, Liabilities, Insurance, Corporations, Taxes, Credit Rating)
* Digital Records of Physical Goods (Real Estate, Vehicles, Personal Effects)
* Virtual Goods (Objects, Gifts, Currencies)

Health Data:
* Health Care Data (Prescriptions, Medical Records, Genetic Code, Medical Device Data Logs)
* Health Insurance Data (Claims, Payments, Coverage)

Other Institutional Data:
* Governmental Data (Legal Names, Records of Birth, Marriage, Divorce, Death, Law Enforcement Records, Military Service)
* Academic Data (Exams, Student Projects, Transcripts, Degrees)
* Employer Data (Reviews, Actions, Promotions)

In addition to this list, there is also the emerging wellness, or "quantified self," data that some users are beginning to collect about themselves through life-tracking companies, including daily or more granular statistics about their bodies and wellness activities. There is travel data including miles, trips and future plans.

‘Service Providers Must Work For the End-User

Most people do not host their own e-mail servers or websites on servers in their basements. Similarly, most individuals will not have the technical skill or desire to actually manage the collection, integration, analysis, permission management and other services needed to derive value from their data. However, the fact that a few users can host their own email means the open standards for email and http are available top to bottom. We what to see Personal Data Services available through open standards, open source code, and an ecosystem that will interact with people who host their own PDS.

But mostly, individuals need to be able to trust that service providers in the Personal Data Ecosystem are working on the user's behalf. Given the sensitivity of the data, and the complexity of running their own servers, most users will rely on Personal Data Service providers. In addition, market models need to emerge that support the Personal Data Store Service Provider making money while working on the users' behalf. The Personal Data Ecosystem Consortium has a Value Network Mapping and Analysis project to outline this model and is raising money to support and foster the model.

Personal Data should be treated like Personal Money

Individuals must be able to move data between service providers, as they can move money between banks, retaining its value. However, with user's data, it's the user that is the provider, but there must still be many takers because of open data formats, activity streaming, and clear identity models that are also portable and separate from the data bank.

End-user choice and the right to transfer data from one service provider to another is key to this model. Just as our money does not become worthless when we move it from one bank to another, the same needs to hold true for individuals’ data.

Consumers need to be able to to Collect and Aggregate Their Data from Product and Service Providers

For this Personal Data Ecosystem and Economy to emerge and for user’s to be properly protected, it is essential that users have easy access to their data from the providers with whom they do business. The steps involved in getting data out of services are tedious and onerous, and often multi-step because we don't have clear "patterns" and open standards for getting data, nor do we require companies to give you a copy of your complete data.

1. Data must be available to users in machine-readable ways using open standards such as Microformats and Activity Streams that are driven by many developers and users, not just a single company. Where data export is available, it is often not machine-readable. Manually exporting repeated monthly statements as they are issued, as a few services offer, is not the answer.

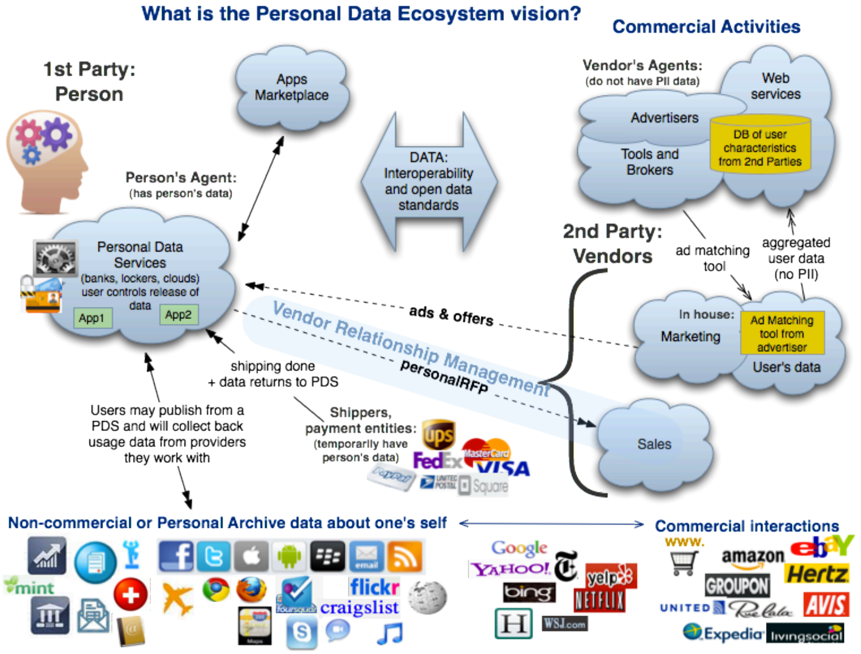
2. Users must have the ability to see and correct their own data, and delete within certain bounds, at sites with which they interact. These tools are not yet created in many cases but we believe with government support, they could be developed and sites that collect data on users could then share that data with users.

3. Simple Internet Open Standards like OAuth allow for personal data stores to link to accounts without the dangerous practice of giving one's username and password to various service providers. Instead, an OAuth token is issued, with username and PW passed only to the issuing party. This keeps users from sharing login information with unscrupulous services and means the OAuth provider doesn't have to "police" a service just to manage login credibility.

4. Portability of data is critical for many reasons, including managing data across providers where businesses fail. People need to be able to move their data to an alternate and hopefully more viable provider in these instances, as well as if they just prefer another provider due to different features and services available. Additionally, to create competition and innovation for Personal Data Services, data must be portable to prevent "lock-in" -- which is currently what many businesses use to prevent users from going elsewhere.

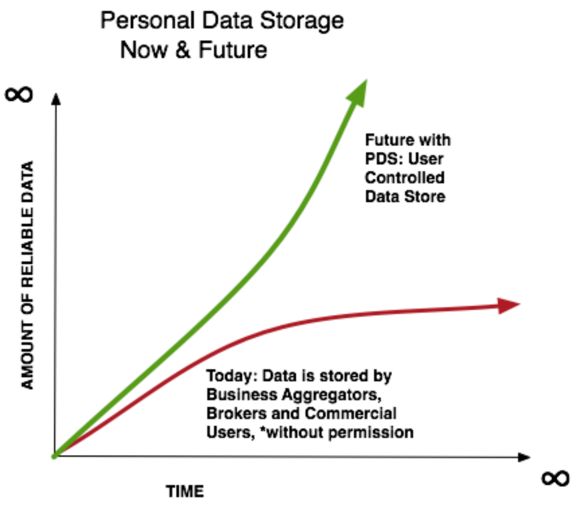
5. Personal data stores and systems must have 4th Amendment protections that require judicial oversight in order for users to feel trust when putting all their data into a single or a few PDSs.

Data transparency, persistence and portability is critical so that as services disappear, user data and digital assets will persist. (For example, the social bookmarking site Del.icio.us makes personal data available to users, and this capability was utilized a lot recently after Yahoo! was reported to be shopping the website). Users create content and generate data during site usage, and those users should be able to easily export their work product from those sites. Business models should not rest on "locked-in" data from users.


What is the Personal Data Ecosystem vision?
Commercial Activities
1st Party: Person
Apps Marketplace
DATA: Interoperability and open data standards
Vendor's Agents: (do not have PII data)
Web services
Advertisers
Tools and Brokers
DB of user characteristics from 2nd Parties
Person's Agent: (has person's data)
Personal Data Services (banks, lockers, clouds) user controls release of data
App1
App2
2nd Party: Vendors
ad matching tool
aggregated user data (no PII)
ads & offers
Vendor Relationship Management
PersonalRFP
In house:
Marketing
Ad Matching tool from advertiser
User's data
shipping done + data returns to PDS
Users may publish from a PDS and will collect back usage data from providers they work with
Shippers, payment entities: (temporarily have person's data)
Sales
Non-commercial or Personal Archive data about one's self
Commercial interactions

Keeping our Data for a Lifetime, If We Want to do so

What if the individual could choose to retain all or a subset of the information about themselves for as long as they wanted? This is a graph that shows today's current data environment and a future where people are in control of their own data, and the opportunities around opt-in, more reliable data than stalking users surreptitiously currently permits.



The red line shows us what's happening today: some data aggregators are necessarily self-regulating by limiting the amount of time they keep data, and governments are limiting data retention and anonymization practices. And much data that is collected is without explicit permission, other than through onerous privacy policy the user agrees to once (usually) and the green line shows us what WOULD happen if people were given the capacity to store and manage their own data – if they could keep as much data as they wanted for as long as they wanted, or not at all, in their own data banks. Digital footprints reflecting a lifetime could be shared with future generations, people could self-assess, and applications through a marketplace would emerge to create new businesses and data uses we haven't yet thought of. In this user-centric model, the individual can aggregate information about themselves, where new classes of services more specific to the individual, based on data accessed with user permission, can emerge.

The foundation of this ecosystem is personal data storage services that are totally under the control of the individual. But a user-centric identity system needs to function in partnership with it (separate from a PDS) and we will need a regulatory regime that supports both of these technology solutions in user-centric form, where users own and control their own data.

This model where individuals are in control of their own data, aligns with the interests of all the stakeholders that we are seeking to balance. Only the data brokers and aggregators lose and they could refactor to have new roles in this ecocsystem of end-user controlled data.

Companies who collect personal data win. By sharing and synchronizing with people's personal data stores, companies get far more accurate information. New services can be offered on data sets, including data not previously permitted to be used or accessed for providing services (telephone log records or mobile geo-location data, for example). And innovation for the PDS and applications marketplace would be a huge new area of development for startups and large companies alike.

People win. By collecting, managing, and authorizing access to their own personal data, users will increase their trust and use of digital realms. This empowers people to work together in communities and groups more efficiently and effectively. Users will be able to see themselves reflected, and participate in transactions more directly with vendors.

Regulators, advocates, and legislators win. By protecting people with new frameworks that also encourage innovation and new business opportunities, government can give people useful tools to interact with agencies because user's identities are trusted.