STARTUP TECHNOLOGY REPORT

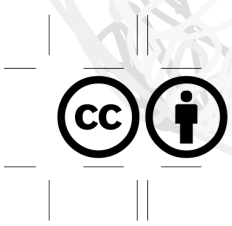
SUMMER 2012

BY: MARKUS SABADELLO

ACQUIRING, STORING, ACCESSING AND MANAGING PERSONAL DATA

PERSONAL DATA ECOSYSTEM CONSORTIUM

[FUNDED BY Internet Society 20 YEARS]



HTTP://PDE.CC/SUMMER-2012-REPORT/

personal data ECOSYSTEM

PERSONAL DATA ECOSYSTEM CONSORTIUM

STARTUP TECHNOLOGY REPORT - PHASE ONE:

ACQUIRING, STORING, ACCESSING AND MANAGING PERSONAL DATA

Markus Sabadello, markus@PersonalDataEcosystem.org
v5, 10 April 2012

This is a report for a qualitative research project conducted by the Personal Data Ecosystem Consortium (PDEC) on technologies used for acquiring, storing, accessing and managing personal data.

Contents

PREFACE ~ LUCY LYNCH

The Internet Society thorough our **Trust and Identity Initiatives** have followed with great interest the work of the user-centric developers and deployers within the Identity EcoSystem. The recent work on Personal Data offers a number of interesting choices for individual users and we are pleased to support this initial survey of some of the leading solutions currently under way. It is our hope this information will encourage additional dialogue within the community and will lead to greater interoperability and better engagement with end-users.

Best -

Lucy Lynch
Trust and Identity Initiatives
Internet Society

PREFACE ~ KALIYA HAMLIN

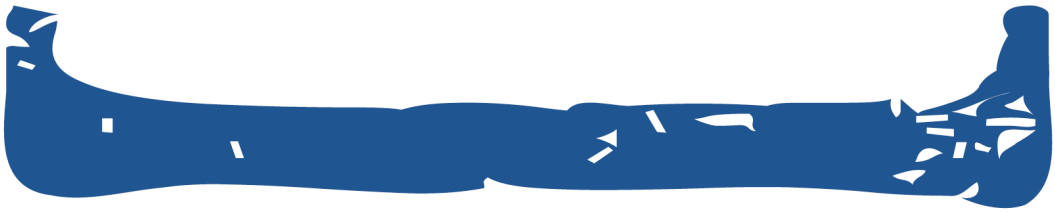
I founded the **Personal Data Ecosystem Consortium** to catalyze a thriving ecosystem where every day people have the power to collect gain insight and if they choose to get value from their personal data.

The PDEC Startup Circle came together a year ago to connect startup entrepreneurs and technologists who share this vision. We successfully fostered shared understanding amongst this growing community, now at over 30 companies. A shared language is starting to emerge; this report is a key part of that process. We're also learning through evidence, discovering the range of approaches that make the vision of a Personal Data Ecosystem a reality.

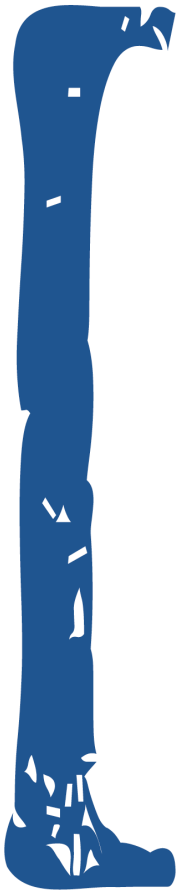
Thanks to the funding from Lucy Lynch at ISOC and the research of Markus Sabadello we are pleased to publish what we hope will be the first of many topical and collaborative reports on key aspects of this emerging industry.

If this report is of interest, consider subscribing to the *Personal Data Journal*, a monthly report we produce about this emerging industry.

Kaliya "Identity Woman" Hamlin
Executive Director
Personal Data Ecosystem Consortium.



ABOUT THE AUTHOR



I, Markus Sabadello, have been active in the Internet identity community since 2007 and have worked as a consultant and developer for several companies that are now members of the Personal Data Ecosystem Consortium. As a participant and observer of many online communities and standard development efforts, as well as of the Internet Identity Workshop and other conferences, I have become familiar with a wide range of paradigms and technologies being used in this space. I run my own open-source initiative "Project Danube", which does not pursue commercial interests and does not strive to deploy production-grade solutions, but which casually experiments with some of the ideas behind user-centric identity, personal data storage and Vendor Relationship Management (VRM). In 2010, it became one of the first projects to achieve interoperability with other codebases of the Federated Social Web community, such as Diaspora and Status.net.

I now contribute to the Personal Data Ecosystem Consortium as a researcher, as well as to its monthly journal as an author and analyst. It is an exciting experience to witness a rapidly growing number of startups and larger companies getting engaged in the emerging Personal Data Ecosystem. While different companies and individuals have very different ideas and approaches, what they all share is they work with great determination on the transformation we are now experiencing when it comes to the management and use of personal data online. I have thoroughly enjoyed every single conversation I have had during this research and have been overwhelmed (in a positive sense) by the many idealistic visions and innovative ideas on how individuals can be empowered, how new economic opportunities can be created, and how our hyper-connected Information Society can be improved. This pluralism of projects is a source of great strength and inspiration, and diversity will be as important to the PDE as biodiversity is to nature.

Interoperability has continuously acted as one of the guidelines for this research. Will different actors within the PDE be able to eventually work together on the technology level? Can a balance be found between interoperability, the diversity of projects, and their individual approaches and objectives? And what will be the process for articulating a common vision of a PDE everyone can subscribe to, and for designing the technologies that can be used to make it real? It is too early to answer these questions, but the research that is hereby presented could be a first glimpse into the similarities and differences that currently exist. I am looking forward to continuing and expanding this effort, and to having many more conversations with the visionaries and entrepreneurs who are actually building this PDE.

INTRODUCTION

The purpose of the Personal Data Ecosystem Consortium (PDEC) is to catalyze a thriving ecosystem that brings value to people and businesses from various efforts working on new visions and software for the management of personal data. PDEC is facilitating information exchange between startups who have strong shared values and similar goal seeking to give individuals more control over their personal data and create new economic opportunities. PDEC launched a monthly Personal Data Journal.

Today, the various companies and projects pursuing these goals are disparate and loosely connected in their efforts. Their common vision is an actual Personal Data Ecosystem (PDE) can emerge, in which different actors can cooperate and interoperate. The scope of PDEC is broad and includes work on various levels (business, legal, technological, etc.). **The objective of this research project is to explore what technologies are currently being used by the various early efforts in the emerging ecosystem, and to create technological profiles along with analyses.**

In advance of this research, some criticism was voiced, in particular regarding the diversity of the different projects considered. It was argued that before attempting to compare apples and pears, it would first be necessary to consolidate visions, terminology and business perspectives, and to agree what the PDE would actually look like. Only then it would make sense to consider the technologies that might be suitable for building it. This is certainly true, and the report presented here is neither meant to predict the nature of an emerging PDE, nor to promote any particular technology.

Another point made in advance of this research was the participating projects should be classified according to their specific goals and their potential role in an actual ecosystem. Also, it was clear from the start this research could only be a small step and a single piece of a puzzle, and that much further work and engagement with related communities and organizations will be necessary.

METHODOLOGY

The research was performed using qualitative interviews with CTOs, lead developers or otherwise technologically knowledgeable representatives from a pool of companies and projects. In this first phase, the pool consisted of the current members of the PDEC Startup Circle, as well as several additional participants that have been identified as being relevant to PDEC's work. A list of research categories and questions has been developed in advance of the interviews to serve as a rough guideline for the research. It was clear from the start that not all research questions would apply equally to all participants. The interviews took about 30 minutes per participant, and typically had a very casual and conversational (but effective) character, rather than strictly working through the sequence of research questions.

For each participant, a profile page was created on the PDEC Wagn (= a structured wiki system). In such cases where a participant was not willing or available to perform an actual interview, research was sometimes based on publicly available information about the participant. Since the work of this community is still rapidly evolving, the interviews were focused not only on the technologies that are being used right now, but also on future plans and ideas mentioned by the participants.

The following is a list of companies and projects that were considered and contacted for the research. The name(s) of contact persons are given in parenthesis. The status column indicates the amount of information that was gathered, and whether an interview actually took place:

Green: An interview was conducted, and all or most information was received.
Yellow: No interview was conducted, but some partial information was received.
Red: No interview was conducted, and no information was received.

Company/Project	Status
Allfiled (Iain Henderson) http://hub.personaldataecosystem.org/Allfiled	Completed
Azigo (Paul Trevithick) http://hub.personaldataecosystem.org/wagn/Azigo	Completed
Buyosphere (Tara Hunt) http://hub.personaldataecosystem.org/wagn/Buyosphere	Not Completed
Connect.me (Drummond Reed) http://hub.personaldataecosystem.org/wagn/Connect_me	Completed
Archify (Walter Palmetshofer, Max Kossatz) http://hub.personaldataecosystem.org/wagn/Archify	Completed
Gluu (Michael Schwartz) http://hub.personaldataecosystem.org/wagn/Gluu	Partially Completed

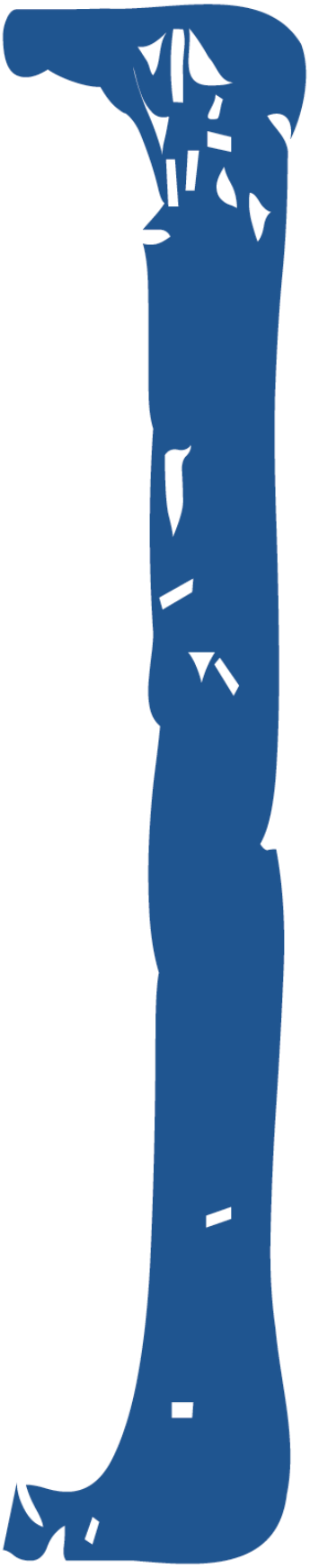
Kynetx (Phil Windley) http://hub.personaldataecosystem.org/wagn/Kynetx	Completed
Mydex (David Alexander) http://hub.personaldataecosystem.org/wagn/Mydex	Completed
MyINFOSAFE (Ross Hughson) http://hub.personaldataecosystem.org/wagn/myINFOSAFE	Completed
Peercraft (Henrik Biering) http://hub.personaldataecosystem.org/wagn/Peercraft	Partially Completed
Personal (Tarik Kurspahic) http://hub.personaldataecosystem.org/wagn/Personal_Inc_	Completed
Privo (Denise Tayloe) http://hub.personaldataecosystem.org/wagn/Privo	Completed
Project Danube (Markus Sabadello) http://hub.personaldataecosystem.org/wagn/Project_Danube	Completed
Qiy (Maarten Louman) http://hub.personaldataecosystem.org/wagn/Qiy	Not Completed
Reputation.com (Owen Tripp)	Not Completed
Singly (Matt Zimmerman, Jeremie Miller) http://hub.personaldataecosystem.org/wagn/Singly	Not Completed
TAS3 (Luk Vervenne, Sampo Kellomäki) http://hub.personaldataecosystem.org/wagn/TAS3	Completed
SwitchBook (Joe Andrieu) http://hub.personaldataecosystem.org/wagn/Switchbook	Completed
The Customer's Voice (Iain Henderson) http://hub.personaldataecosystem.org/wagn/The_Customer_s_Voice	Completed
Tangled Web Communications (Ankit Kapasi) http://hub.personaldataecosystem.org/wagn/Tangled_Web_Communications	Completed
Weqaya	Not Completed
Pidder (Stefan Lipgens) http://hub.personaldataecosystem.org/wagn/Pidder	Completed
Personal Information Brokerage (John Harrison)	Partially Completed

Privowny (Hervé Le Jouan) http://hub.personaldataecosystem.org/wagn/Privowny	Completed
Status.net (Evan Prodromou) http://hub.personaldataecosystem.org/wagn/Status_net	Partially Completed
TheWriteID (Tim De Coninck) http://hub.personaldataecosystem.org/wagn/TheWriteID	Completed
VDesk	Not Completed
Dropbox	Not Completed
Greplin (Daniel Gross) http://hub.personaldataecosystem.org/wagn/Greplin	Not Completed
Continuum Labs (Bill Nelson) http://hub.personaldataecosystem.org/wagn/Continuum_Labs	Partially Completed
Unhosted http://hub.personaldataecosystem.org/wagn/Unhosted	Partially Completed

ROLES AND OBJECTIVES

As has already been mentioned, the participants are rather diverse and different from one another, which makes it hard to directly compare them. While some focus on providing a centralized and universal “Personal Data Store” where individuals can store all their personal data for a wide range of purposes, others aim at providing rather specialized tools (such as mobile apps) to empower individuals, and yet others try to design complete architectures for a large number of different actors and use cases.

As a consequence, one of the first questions that was always asked during the interviews was about the objectives that projects set for themselves. These answers have been collected on the PDEC Wagn system (see the links in the above table).

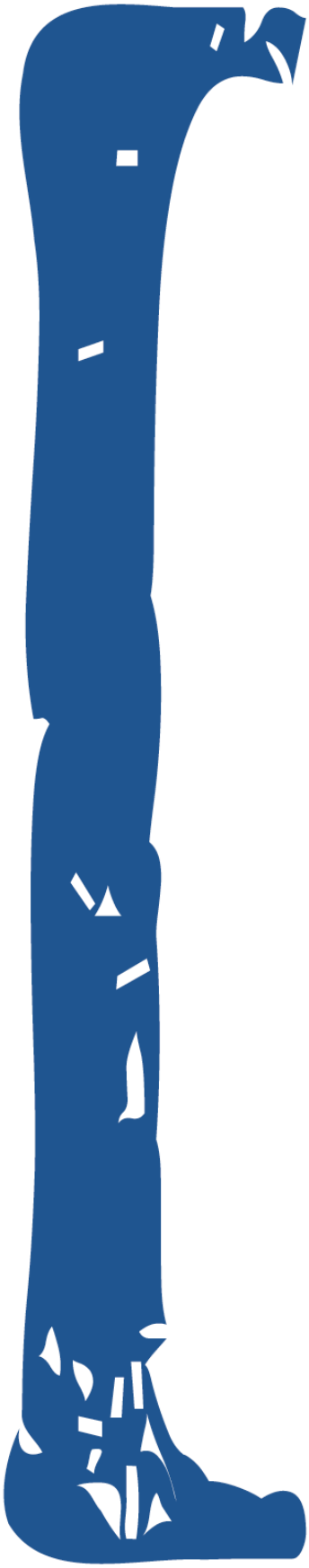


TERMINOLOGY

While this was not a primary focus of the research, it has become clear that exploring and establishing a common terminology would be an important effort for the emerging PDE. For example, terms such as “Personal Data Store”, “Personal Data Locker” or “Personal Data Vault” are used by different companies to describe roughly the same concept. On the hand, there are also variations of a single term that might be irritating. For example, “Personal Data Store” might be considered by one company to only denote a user’s personal data that is stored within a core service provider, while another company might use the same term to describe a central place where a user can manage all of their personal data, regardless of where exactly it is actually stored.

The PDEC Wagn could be a suitable instrument for collecting and maintaining different terms and definitions, e.g.:

- http://hub.personaldataecosystem.org/wagn/Personal_Data_Store
- http://hub.personaldataecosystem.org/wagn/Personal_Data_Vault
- http://hub.personaldataecosystem.org/wagn/Personal_Data_Service



ARCHITECTURES

The participating projects take very diverse architectural approaches when it comes to the storage and transmission of personal data. While some of them are centralized cloud-based services, others place a focus on federation, peer-to-peer and self-hosting patterns. These different approaches can of course be directly traced back to the different roles and objectives the services set for themselves.

Most systems that are based on cloud-based services also offer client software such as browser plugins or mobile apps (or are planning to do so) for accessing and managing personal data. In such cases, the client-side apps typically fulfill one of two purposes (or both): They can allow users to access and manage the personal data in the cloud-based service. Or they can collect personal data from the user and upload it into the cloud-based service.

Some systems put an emphasis on the fact that personal data storage is clearly separated from applications that use it, for example the LifeDash platform, the Unhosted project, or the Kynetx Rules Language (KRL). Some systems that provide a “default” backend service for data storage also plan to offer open source versions of the backend which adept user can set up and host themselves, for example in the case of Azigo or SwitchBook. Another approach is to offer integration with external data hosting services such as Dropbox, for example in the case of Archify. And yet others do explicitly point out that they do not want to offer yet another place where users have to sign up and store personal data, but rather provide tools for managing the accounts they already have online, for example TheWriteID.

The following table tries to roughly classify projects by their architectural patterns:

Centralized Deployment by Default	Focus on Federation	Client-Only
The following systems emphasize the possibility for individuals to sign up at a centralized cloud-based service, but they may alternatively also allow users to switch to different service providers, or even host software themselves. • Azigo • Allfiled • Connect.me • Archify • Kynetx • Mydex • Personal • Pidder • Privo • Privowny • SwitchBook • The Customer's Voice • TheWriteID	The following systems are designed for being deployed at multiple locations, by different services providers or actors in the ecosystem, or even hosted by individuals themselves. A focus is placed on federation, i.e. the ability for different deployments or different parts of the system to interact with each other. • Status.net • Gluu • Project Danube • TAS3	The following systems are aimed at empowering individuals in their use of computers and mobile devices, without depending on a cloud-based service. • myINFOSAFE (Personal data is stored on a user's local PC) • Tangled Web Communications (Personal data is stored on mobile devices and can be shared via the Person2Person peer-to-peer technology)

IDENTITY AND AUTHENTICATION

Different projects also have very different approaches towards identity, i.e. the way in which they conceptualize and authenticate users. This can range from classic username/password identification to models where no user identifier is needed at all. Another interesting aspect of identity is whether projects support identifiers with a scope beyond their own internal needs, and whether they support discovery, identity federation, and other extended functionality.

Several systems that currently use passwords (or passphrases) as primary credentials also support (or are planning to support) multi-factor authentication, e.g. biometrics in the case of Mydex, or Personal, which is planning to implement closed-loop authentication with e-mail or SMS, or with enrolled mobile devices that are registered to the individual. Pidder also supports a wide range of advanced authentication features, e.g. private keys, one-time passphrases, a virtual keyboard, and an authentication puzzle on the screen.

In some cases, signing into the system is possible with different sets of permissions (or clearance levels), e.g. with Pidder.

Systems that issue their own identifiers (e.g. usernames) can also potentially act as Identity Providers (IdP) for others, e.g. by using the OpenID Connect or SAML2 protocols. While in some cases, this functionality is consciously left out in order to concentrate on the core service and not confuse users (e.g. Azigo, Personal), in other cases supporting IdP functionality for identity federation scenarios is a key part of the overall vision (e.g. Gluu, Privo).

Most systems do not position themselves when it comes to a requirement of using real names, i.e. it is possible (and allowed) to use them with pseudonyms. In the case of Connect.me, the policy is to not require the use of real names, but to only allow a single account per individual. In the case of Privo, the core service is to actually verify a parent's real name (in order to achieve COPPA compliance – permitting the data of their children to be collected).

All systems where individuals register for an account also support (or are envisioning to support) a "right to be forgotten" in the form of complete deletion of their account.

The following table summarizes identity aspects of the various different systems.

Company / Project	Identity and Authentication	Comments
Allfiled	Username/Password/PIN	IdP and RP roles are supported.
Azigo	Username/Password	E-mail addresses such as yourname@azigo.com are also created, which can be used to establish a communication channel for the e-commerce relationships of an individual
Connect.me	Social Sign-In	A username has to be chosen, which becomes part of the individual's card URI, e.g. http://connect.me/yourname
Archify	Username/Password or Social Sign-In	

Company / Project	Identity and Authentication	Comments
Gluu	Identity “virtualization” of existing organizational data stores, such as LDAP or Active Directory	Focus on SAML2 for identity federation.
Kynetx	Username/Password or Social Sign-In	
Mydex	Username/Password, also multi-factor authentication supported	SAML2 and OpenID Connect IdP role is supported.
MyINFOSAFE	Local Username/Password	
Personal	E-Mail/Password	
Pidder	Username/Passphrase, also: Private key file, one-time passphrases, solving of an authentication puzzle on the screen, virtual keyboard on the screen to avoid keyloggers. Brute-force protection is also built into the website.	Signing in can be done at a certain clearance level, which means that during the session only operations at this level will be possible and data of higher confidentiality will not be delivered in the first place. This could be useful e.g. in Internet Cafés, where one would not want to sign in to one’s service with full privileges. Role as OpenID provider is on the roadmap. A “Quicklogin” feature allows one-click authentication to other web based services.
Privo	Username/Password	A “Privo Connect” service similar to “Facebook Connect” is planned. E-Mail addresses are used to look up individuals. An E-Mail address can be shared by parents and their children.
Project Danube	I-Name/Password	
TAS3	Many different identification/ authentication methods are supported by the architecture, e.g. Yubikey or one-time passwords.	SAML2 and OpenID Connect IdP and RP roles are supported. Pairwise Pseudonymous Identifiers (PPID) are supported at all layers. Veronymity is supported upfront through configuration, or in case of abuse (following a court order to IdP).
SwitchBook	No identification and authentication of individuals, but unique identifiers for local files containing personal data.	
The Customer’s Voice	Username/Password/PIN, additional approaches in research.	IdP role is supported.

Company / Project	Identity and Authentication	Comments
Tangled Web Communications	No explicit identification and authentication needed, since personal data is only stored on the mobile device.	
Privowny	E-Mail/Password	Anonymous e-mail addresses such as 6jp1897g@my.privowny.com can be created, which can be used similar personas to give companies only access to a certain subset of an individual's personal data.
Status.net	Username/Password	Webfinger-compatible identifiers are issued for discovery purposes.
TheWriteID	Username/Password	Social Sign-In to social networks for the purpose of managing personal data in those networks.

DATA MODELS AND NATURE OF PERSONAL DATA

One obviously important aspect of a PDE (and perhaps the aspect that is most relevant for interoperability) is what kind of personal data is stored, how it is modeled, how it is expressed in internal storage systems, and what schemas and semantics are used to describe and structure it. In this area also, approaches of the participants vary greatly. While some work with traditional models such as relational databases, XML files, others use NoSQL-based storage mechanisms or semantic RDF stores. In most cases, the data model for the stored personal data is extensible.

Some systems place a focus on acquiring personal data automatically (e.g. by screen scraping, by capturing data from web forms, by importing data from social networks, or by capturing your browsing history). Examples include Azigo, Archify, Privowny, Switchbook and Singly. In such cases, the key idea is to offer value-added services to the user based on the recorded data. Some systems take the opposite approach of emphasizing a user interface (web or local) for adding/changing/removing personal data manually, which can then be administered, shared, etc.

The following table provides an overview of data storage mechanisms, data models, and the kinds of personal data that are stored and managed by the various systems:

Company / Project	**Data Storage / Data Model**	**Nature of Personal Data**
Allfiled	Combination of relational and RDF database	900 attributes in a rich, deep person-centric data architecture
Azigo	RDF Quads-based model consisting of three layers: 1. Context data model, 2. Higgins data model, 3. Persona ontology (re-using existing vocabularies such as vCard, FOAF, schema.org)	Everything related to your e-commerce relationships, e.g. shipping/billing information, web form contents, interests, brands you like, etc.
Connect.me	NoSQL database	Social graph, in- and out-going vouches, reputation graph
Archify	MySQL for core user data, Redis for various structured data, Lucene ElasticSearch for full-text search	Everything that acts as your sources of information online, e.g. websites you browse, social networking activities, e-mails.
Gluu	OpenDS/OpenDJ LDAP server	
Kynetx	MongoDB with no specific schema.	Kind of data depends on rulesets.
Mydex	RDF data store with an extensive “master data schema” which integrates with a range of external ontology and schema. Includes thousands of attributes and can be extended.	Personal information, transaction history, preferences, intentions, browsing history, address books, credentials, proofs of claim, etc.

MyINFOSAFE	XML files with XML schema	5 sections: My Information, My Life, My Finances, My Assets, My Health
Personal.com	HBase (Hadoop). Schema is defined in XML and extensible. Data fields are organized in “gems” for specific purposes.	Any personal data used in online interactions.
Pidder	Relational database. Personal data can be organized in “segments”, “cards”, “wallets”, “personas”, and identities”.	Any personal data used in online interactions.
Privo	MySQL	Relationships between parents and children, plus online contact data (e.g. e-mail addresses)
Project Danube	Native XDI data store based on Berkeley DB.	Basic profile data, plus social networking data such as list of contacts, status updates.
TAS3	Agnostic to any specific storage or model. Data interoperability is handled by ontologies and mappings.	Agnostic to any specific kind of personal data.
SwitchBook	RDF data serialized to files (“portable contexts”) in Turtle format	Various behavioral data of activity tracked on the web, as well as organizational data (e.g. folders, annotations).
The Customer’s Voice	Hybrid model consisting of a relational store, semantic store, and raw store. Schema developed based on extensive experience with CRM systems, about 4,500 attributes.	Rich, deep schema designed to cover most of the personal data needs an individual encounters in their online buying.
Tangled Web Communications	SQLite files on mobile devices. A “master data model” is being developed that can be shared by different mobile apps.	Any kind of data needed by mobile apps.
Privowny	Relational database.	Everything that companies know about you, including data you submit in web forms, preferences on web sites, and data from companies’ cookies.
Status.net	Relational database.	Various personal data, mostly focused on social networking needs, e.g. status updates, personal profile, uploaded photos, comments, etc.

TheWriteID	NoSQL database. Data model is developed based on data models of the various networks you use. “Variables” can be used in cases when you use different data in different network.	Profile information used by various social betworks.

DATA ACCESS

All systems that offer individuals the ability to store personal data also offer (or are envisioning to offer) one or more APIs that make it possible to share personal data with 3rd parties. Sometimes, a distinction is made between “private APIs” (which are only intended for use by client software from the same company/project) and a “public API” which is open to anyone wishing to interface with the system.

Many systems design their API according to widely accepted Web 2.0 principles, e.g. RESTful JSON based APIs. Some systems that put a focus on semantic data models (e.g. RDF, XDI) also offer appropriate API endpoints (e.g. SPARQL, XDI Messaging). Yet others support classic SOAP web services, or access to personal data attributes via identity federation protocols such as OpenID Connect or SAML2.

The following table summarizes planned and actual available APIs to access personal data within various systems:

Systems with RESTful APIs
- Connect.me (RESTful/JSON/OAuth 2.0) - Archify (RESTful/JSON/OAuth 2.0) - Personal (RESTful/JSON/OAuth 2.0) - Mydex (REST/JSON/XML/OAuth 2.0) - TAS3 (RESTful/OAuth 2.0/UMA) - The Customer’s Voice (RESTful/OAuth 2.0/UMA) - Allfiled(RESTful/OAuth 2.0)
Systems with Semantic APIs
- Azigo (SPARQL) - Switchbook (SPARQL) - Mydex (SPARQL) - Gluu (XDI Messaging) - Project Danube (XDI Messaging)
Attributes via Identity Federation
- Gluu (SAML2) - TAS3 (SAML2, OpenID Connect, ID-WSF)
Others
- TAS3 (SOAP CRUD API) - Tangled Web Communications (Personal data is stored exclusively on mobile devices and can be shared via SMS) - Status.net (OStatus for Federated Social Web) - Project Danube (OStatus for Federated Social Web) - Kynetx (Evented APIs specification based on event-driven interactions rather than traditional request/response calls)

Privacy and Security

Due to the sensitive nature of personal data, many participants have placed a special emphasis on privacy and security (or are planning to do so). This includes the actual storage of personal data, the various ways in which it can be accessed, managed and shared, and the policies and permissions associated with such access. “Privacy by Design” – the idea of building a system’s architecture and code in an inherently privacy-protecting way – is a keyword that was mentioning by several participants.

All systems that provide private or public APIs allow (or even require) the use of HTTPS, to ensure confidentiality of the personal data that is transmitted over the wire. When it comes to authorizing access to APIs, OAuth 2.0 seems to be the mechanism of choice for many systems. User Managed Access (UMA) is also in use.

Some systems encrypt personal data they store on a local device or in the cloud. This serves to keep it private not only when internal storage of a service provider is compromised, but also to keep it hidden from the service provider itself, for a maximum of privacy. To achieve this, users are issued private keys which are not accessible by the service provider. The primary challenge here is key recovery, i.e. to cater for the case when users lose their key. Another challenge is that encrypted personal data is not searchable. Different strategies exist regarding which data should be encrypted. All data may be encrypted by default (e.g. myINFOSAFE), or only certain “sensitive” data as considered by applicable legislation and guidelines) may be encrypted (e.g. Personal), or encryption may be manually turned on and off for certain data (e.g. Privowny). For maximum privacy and security, the private key may be protected by a security code or split into multiple parts and stored in different locations in the cloud.

Pidder achieves reliable cryptographic functions through the use of a special-purpose browser plugin (“Cipherbox”).

In the case of Mydex, access to APIs is only given to 3rd parties that have been explicitly verified and certified.

Some systems actively clean server log files from personal data that might have been received via an API or otherwise processed, e.g. Personal.

Some systems that acquire personal data automatically while browsing the web support ways to disable their functionality in certain situations to improve privacy, e.g. by blacklisting individuals domains, or by using a browser’s incognito mode (e.g. Archify).

INTEROPERABILITY AND OUTLOOK

As has been mentioned in the introduction, a small research project like this can only be a small step and a single piece of a puzzle. For developing a vision of the future nature of the PDE, much further research is needed, as well as effective discussions and consensus finding processes. Based on the experiences and results of this initial research, a second phase could be conducted in which the research questions would be applied to a larger audience. This second phase should be based upon a more formal research plan and survey instrument, and could include the use of an automated survey tool such as SurveyMonkey.

There are several challenges on the way towards establishing an actual PDE, for example the joint development of legal frameworks and business models, and a much more concrete agreement of the nature of the PDE. On the technological level, interoperability between different efforts appears to be the long-term vision on the horizon, given the fact that individuals' control over personal data can only be fully realized if they enjoy data portability, i.e. the ability to choose from as well as move between different service providers. The question of how to achieve this is obviously a politically sensitive topic, and it is too early to realistically predict at this time which technologies could eventually serve as the foundation for a common PDE in which different companies and code bases can interoperate seamlessly.

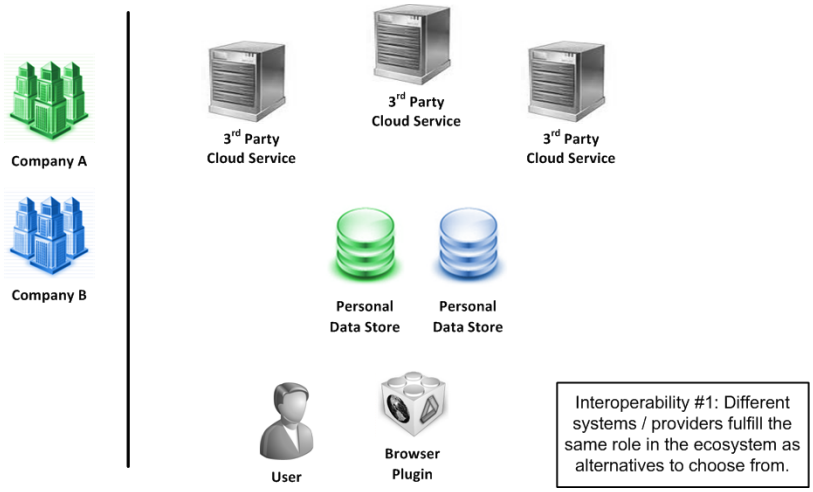
Conceptually, there seem to be two broad ways in which any two projects can interoperate with each other, depending on whether they fulfill roughly the same role or not. It appears that for a fully pluralistic PDE to become reality, both types of interoperability are necessary.

Interoperability Type #1: Same Roles

Consider for example two similar systems A and B. They both offer a “Personal Data Store” service, where individuals can register for an account. They can store and manage all their personal data online, and both A and B provide similar ways for individuals to access and share their personal data, e.g. with companies that are interested in subscribing to it.

In this scenario, interoperability can be achieved if there is a way for individuals to switch from A to B, or vice versa. Data portability must be ensured, i.e. it must be possible to seamlessly move personal data from one service to the other, just like money can be moved from one bank to another. Also, there should be a continuity of services, e.g. a subscription to an individual’s “Personal Data Store” should still work (or be easy to re-establish) after the move is complete.

To achieve this kind of interoperability, different systems would have to agree on a number of technical details, for example the internal data model they use for expressing personal data, and the exact mechanisms for moving personal data. Several systems today already provide mechanisms for exporting all stored personal data (e.g. Azigo, The Customer’s Voice, Tangled Web Communications, SwitchBook, TAS3, Project Danube, Personal, myINFOSAFE). Such data could then theoretically be imported into a different system.



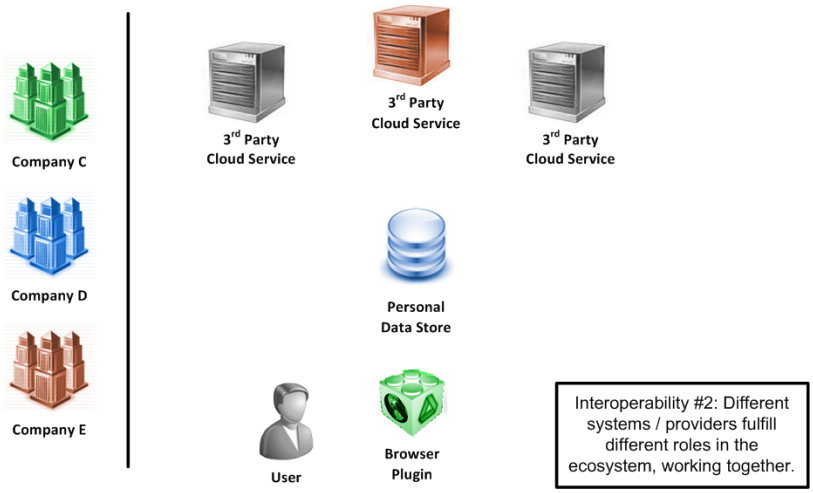
Note: This diagram and the one on the next page don’t imply a PDE must follow a specific architecture or must include certain components. A PDE can also look very different. These diagrams only illustrate different types of interoperability between systems.

Interoperability Scenario #2: Different Roles

When considering systems that fulfill different roles, then interoperability refers not to mechanisms for moving from one system to the other, but to the ability to use them at the same time, as complements. Consider for example project C to be a browser plugin that augments the browsing experience in a personalized way, project D to be a “Personal Data Store” service, and project E to be a marketplace service where companies can discover and subscribe to individuals’ personal data.

In this scenario, interoperability can be achieved if C, D and E work together, rather than being alternatives to each other. The browser plugin C will use the APIs of the “Personal Data Store” D to provide the personalized experience. In addition, it can feed new personal data back into D. Through the marketplace service E, companies can subscribe to the “Personal Data Store” D. And the browser plugin C can in turn display new offers received via the marketplace service E.

Many examples of how services with different roles could complement each other are possible. To achieve this kind of interoperability, agreement is needed not so much about the inner workings of services, but about the APIs, standards and protocols they use to communicate with each other.



During this research, participants were always asked two questions: “Does your system work in any way with another system in the PDE?” and “How COULD your system work with another system in the PDE?”

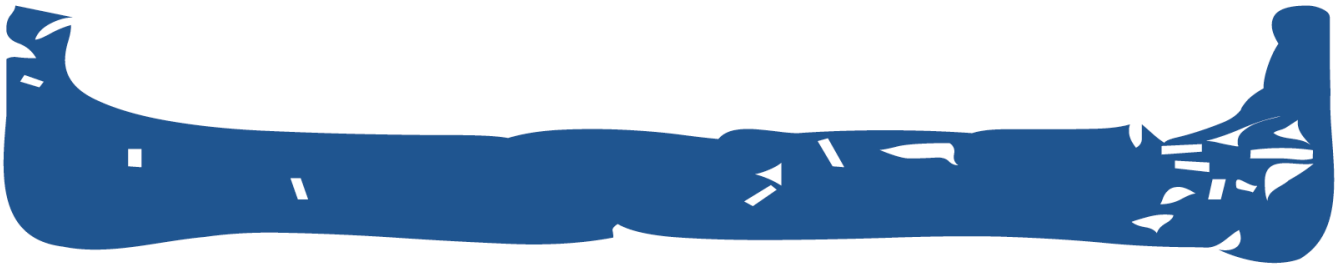
Virtually all participants of this research stated that they consider interoperability an important goal. Only a few companies/projects, however, seem to have resources at this time to work on realizing it.

Some limited interoperability can already be achieved today in cases where systems support common protocols (e.g. SAML2 for identity federation). Some have concrete ideas on how their system could be made to work with another system, and some have very well developed, extensive visions of how a bigger ecosystem with multiple actors and use cases could look like (e.g. the TAS3 architecture, or Connect.me’s “Respect Network”).

Over time, concrete cases of different systems working together in the management and exchange of personal data should be collected and documented. At some point, similar to the approach that was taken by the Federated Social Web effort, PDEC could define a “SWAT0” test with a very basic interoperability scenario that should consist of a general idea, but not mandate any concrete technology. Example ideas include:

- Export data from your system into a different system within the PDE.
- Authenticate to your system using a different system within the PDE.
- Use your system’s user interface to access data from a different system of the Startup Circle.

At a time when consensus emerges within the ecosystem for specific technologies (protocols, formats, etc.), a suite of very concrete and detailed interoperability tests could be developed, potentially with a reference implementation to test against. This approach would be similar to what OSIS has done in the identity world (e.g. OpenID or InfoCard interoperability testing).



FUNDED BY

Internet Society 20 YEARS

HTTP://PDE.CC/SUMMER-2012-REPORT/