Internet Privacy Workshop Position Paper

Personal Data Service

Paul Trevithick, paul@azigo.com

Privacy can be framed as the requirement to give the user with control over whom can see what aspects of that user's personal data. The question is how to deliver this control when single person's information is stored in thousands of databases physically distributed across the Internet, is stored under a wide variety of policies including access control rights, is described using few common data models, and cannot be accessed using common protocols. We describe a software project called Higgins [1] within the Eclipse Foundation that seeks to answer to this question by developing a user-controlled, cloud-based service called a *Personal Data Service (PDS).*

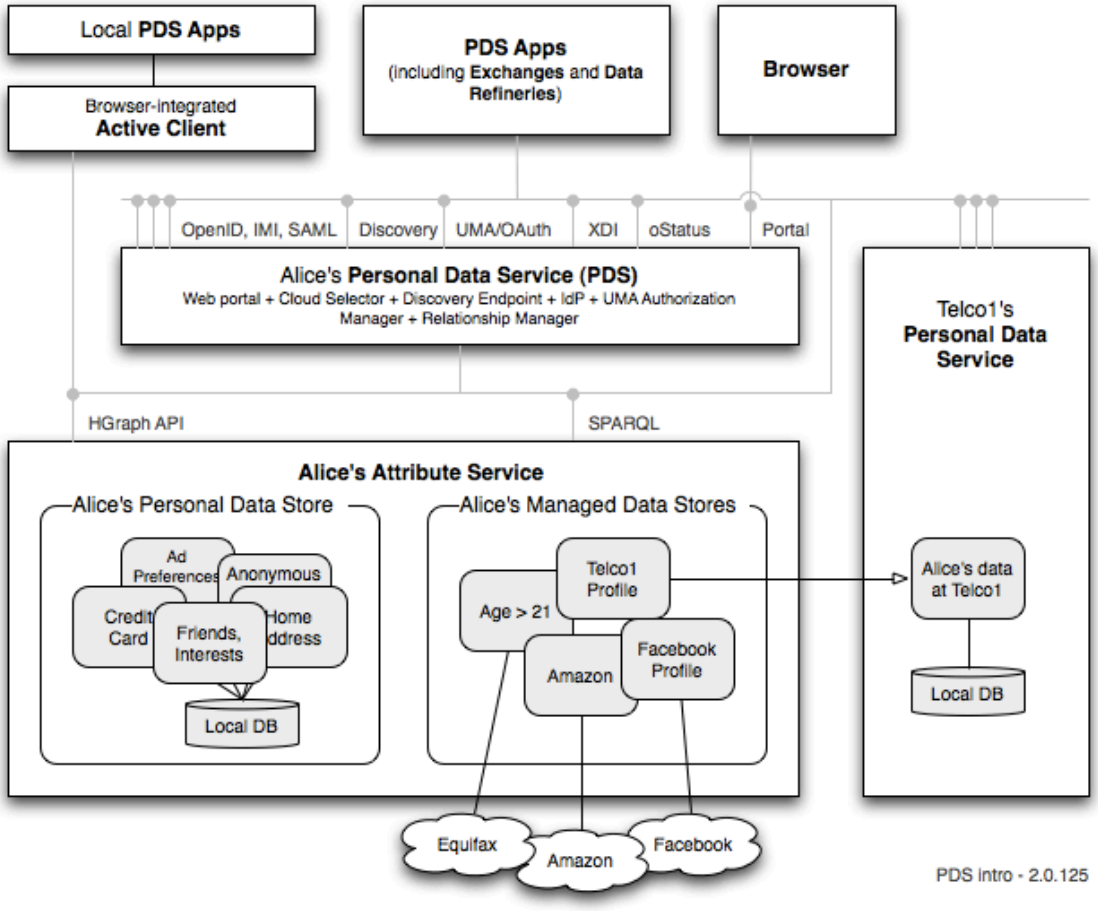
Today a person’s digital identity, (including attributes and social graph), is stored and managed by data silos that are at best only partially under the user’s control. The user isn’t even aware of many of these silos, and in most cases their data is subject to terms of service and privacy policies that may change at any time. You can search the Web for information about a person, but you can’t ask a person for their digital identity—they have no way to manage or store it. The core thesis behind a PDS is that the best way a person can exert maximal control over their own personal data in external services (e.g. Facebook, data brokers, government databases, etc.) is to have a data service of their own; one that works on their behalf.

In a sense, a PDS extends the traditional concept of a *user agent* to allow it to (i) manage a rich, declarative digital description of the user linked to multiple external sources (ii) put the disclosure of this information to apps and services under the direct control of the user. Another perspective is that a PDS introduces a missing identity layer sandwiched between applications “above” and operating systems “below”.

Control. A PDS provides a central point of control for information about a person, including their interests, affiliations, friends, and so on. The PDS is a place where the user can control data flows between services that provide data about them, and services that wish to consume it. In some cases the data itself flows directly between the data source and the data consumer, while in others the data flows through the PDS as an intermediary. In some cases the PDS is the originating source of the data.

Data Management. In cases where data flows from or through the PDS, we have the opportunity to map it into a normalized data model, provide the ability to see the data values, and in some cases be able to edit and update it.

Discovery. A PDS supports a discovery API that allows the user to be discoverable by other people, organizations, apps and exchanges when the incoming inquiries meet criteria the user specifies.



Interoperability. Each PDS is a peer that can exchange personal data with other PDS peers within a distributed network operated by a multiple organizations. Each PDS would be hosted by a trusted organization that acts on behalf of the individual, or be would be self-hosted.

Status

The Higgins PDS, while only partially complete, has been incorporated into commercial pilot projects and is evolving from this experience (e.g. new attribute types are constantly being added to its underlying RDF/OWL-based data model [2]).

[1] http://wiki.eclipse.org/Personal_Data_Service_Overview
[2] http://wiki.eclipse.org/Persona_Data_Model_2.0