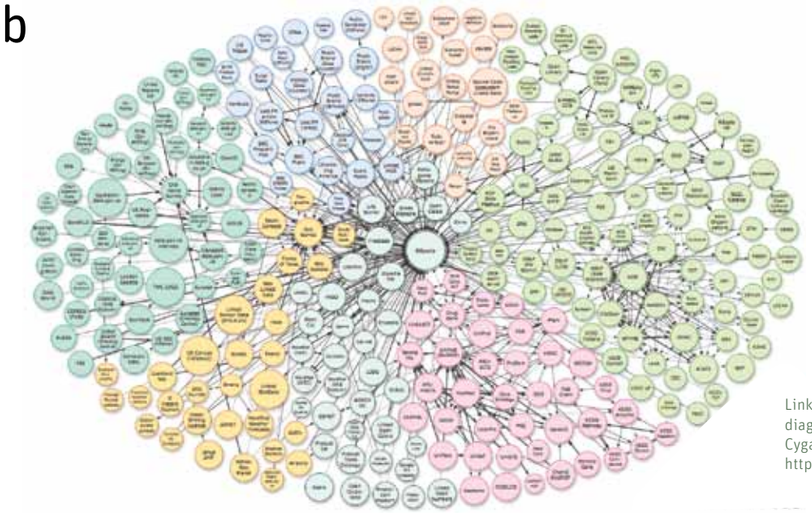
Authentication and Authorization
in the Semantic Web

Life-long learning requires new concepts for storing and retrieving data of curricula vitae. In a project, BFH-TI has explored the possibilities the semantic web has to offer, especially focusing on authentication and data integrity



Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/

A bit of history

In a 2001 Scientific American article[1], Tim Berners Lee (best known as the inventor of the World Wide Web) and others published a vision of a «new» web of data, a semantic web.

Addressing issues with the current web, namely its lack of semantics[I] , the semantic web should create the possibility for machines to interpret the data represented and derive (even additional) meaning from it. To illustrate this, let us have a look at a simple but common example: Searching the web for a term – for instance «jaguar» – will lead to a variety of results, including information about the animal, the car as well as the version of the MacOS-X operating system sharing the same name. How should a machine be able to know which instance we are looking for?

One goal (besides others) of the semantic web is to resolve issues like this by providing ways of semantically interlinking data. In the past ten years the ecosystem around this idea has grown significantly and a vast landscape of standards, protocols, tools as well as semantic content has emerged. Of course, old and new challenges have surfaced, one being that of authentication and authorization[II] in linked data[III]. Research at the intersection of identity and access management as well as semantic web technologies fits perfectly into the portfolio of ICTM's IAM research group.

In a first project, we used semantic web technologies to represent the contents of a curriculum vitae and to implement access control to the service publishing it. This initial research led to promising results but also showed open issues. Some of these will now be researched more thoroughly in a follow-up project.

The CV3.0 Project

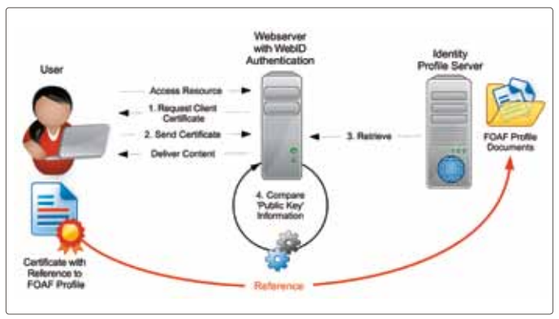
In BFH's CV3.0 project, the contents of a curriculum vitae are to be represented as linked data. This data needs to be digitally signed in order to be properly verifiable by a third party and access to it must be authenticated. In cooperation with the department of Business Information Technology of BFH we have developed a set of proof-of-concept implementations, covering various technological aspects of such a future «CV 3.0» as well as accompanying methodologies and documentation needed to complement the technical parts.

Besides having data in the form of files out of a common institutional document management system (like a bachelor diploma for instance), the main content of the CV is represented as linked data, served over a HTTP-based interface. Access to this data needs to be protected, thus an authentication mechanism is absolutely required.

The project team has chosen WebID, rooted in the semantic web, for authentication and has done thorough research on authorization standards atop of it.

WebID – performing authentication with linked data

In 2008, a specification initially called FOAF[IV]+SSL emerged from work done by Henry Story and others. Incubated by the W3C[V] , the protocol was renamed to WebID and is now further developed by a W3C community group. The main idea of WebID is to perform a secure, encrypted authentication using certificates mapped to information found elsewhere in a so-called FOAF profile document. This document has to be reachable for the server handling the authentication request and contains information uniquely identifying the given certificate. Optionally, it can contain additional information about the user. The basic steps are as follows:



Authentication with WebID, showing the different steps including the retrieval of the FOAF-profile from the profile-server.

1. Upon connection to a WebID-enabled service, a certificate is requested from the connecting client.
2. The certificate delivered by the client contains an additional field with a reference to its FOAF profile document.
3. The server retrieves this document and compares the information concerning the cryptographic key of the client's certificate with the certificate given by the client.
4. If the information matches, access is granted; other wise access is denied.

Advantages of the WebID protocol include the use of standardized and widely used technologies for certificates and encryption, decentralization and availability to a wide range of clients (not only restricted to web browsers).

During the research for the CV3.0 project, we determined that there is no commonly available software (a so-called identity provider) which could be used to generate WebIDs for a whole institution or enterprise in an easy way. Thus we decided to create our own implementation, WebIDP. WebIDP allows the user to log-in using their normal company credentials (like the BFH-login in our case) and then generate one or more WebID(s) to use with different browsers/devices. Also, generated WebIDs can be managed and revoked.

Further research then investigated how to use the created WebIDs to protect the access to the various kinds of CV3.0 data, by using special languages to model authorization restrictions.

Outlook – continuing our work with PerSemID

The work done in the CV3.0 project serves as a great starting point for diving deeper into various aspects of authentication and authorization in the context of linked data. It has shown, especially regarding authorization, that there are currently many open questions waiting to be solved.

In PerSemID, our next BFH research project (again in cooperation with the department of Business Information Technology), which started in February 2014, we plan to tackle some of these questions and intend to interlink semantic identities with the concept and infrastructure of the Swiss national electronic identity (SuisseID).

Authors:
Pascal Mainini
Research Associate
Prof. Gerhard Hassenstein
Lecturer in IT Security

Kontakt
– pascal.mainini@bfh.ch
gerhard.hassenstein@bfh.ch
– Info: http://cv3.bfh.ch

Glossary

I Semantics: in this context, «meaning» of the data by describing relations and what they stand for.
II Authentication: identifying something/someone. Authorization: after identification, decide to what something/someone has access and how (read/write for instance).
III Linked data is semantically interlinked data represented in RDF.
IV Friend of a friend, a popular ontology for representing personal data as well as one's social network.
V The World Wide Web Consortium, responsible for many web-related standards.

[1] http://www.scientificamerican.com/article/the-semantic-web