Abstract. In the coming years we will see a revolution of machine knowledge and abilities which will emerge from different activities and trends in three distinct areas connected to the internet: the emerging Web of Data and Web of Services, as well as the Web of Identities, which we define in this paper. These areas are about making accessible, connectable and processable semantic knowledge about data, functions and individuals, respectively. We outline current activities and trends and point out the impact enabled by each of them. After identifying visionary scenarios exploiting the blended knowledge of these activities, we highlight the impact of the webs on the future of the internet.

1 Introduction

The amount of available information is growing exponentially. Sources are developed or made accessible, content is produced en masse through the paradigm shift of consumers becoming producers in the Web 2.0, users leave data about themselves and their social connections, companies are unlocking services.

Every day it becomes more complex and difficult to make this information findable and usable. In the near future, e.g., a search engine’s one dimensional list of ten results returned on the first page for a query is not sufficient anymore to cover our needs. Users will demand to ask more sophisticated, natural queries. To answer those, today’s search engines’ primitive semantic understanding of the content they index is not sufficient as a knowledge base.

It is indispensable that machines are taken to another level of understanding: Understanding what terms are about and semantically connected to. Understanding what they can do by detecting services, understanding what they do and successively invoking them to generate further or process existing information. And finally, understanding the user. That is both, understanding an acute need like a query as well as having access to profile information, given an access authorization.

We sketch an answer to the challenge of providing tomorrow’s machines with a toolbox to find and interpret knowledge, to discover, orchestrate, and invoke services with the knowledge gained to solve highly complex tasks.

In the next section we discuss selected activities in the areas of the Web of Data, the Web of Services and the Web of Identities and illustrate how they interrelate and form an emerging big picture of the webs for machines.
2 The Webs

Current research activities focus on making accessible semantic knowledge from open data (Web of Data) and making accessible semantically annotated services (Web of Services) to machines. Emerging from the Social Web [1], we want to introduce the Web of Identities. This web is about users, their assets, facts, preferences, social graphs etc. This content is highly privacy-sensitive, volatile and valuable. Different user-centric access control mechanism will be needed than those of the Web of Data.

In the following we want to briefly introduce and explain all three Webs. It must be noted that we do not include another often mentioned Web, the Web of Things. In the view that we sketch here, devices and the data that they can provide are orthogonal and therefore either part of the Webs we are about to introduce, or out of scope of this paper.

2.1 Web of Data

Definition 1 (Web of Data). The Web of Data is a distributed web of interconnected data sets of semantically annotated data.

Fig. 1. Web of Data: A distributed web of semantic interconnected data sets containing general knowledge. The data sets consist of a triple store which contains structured knowledge and which may link to knowledge contained in another data set.

Problem Motivation The idea of the Web of Data originated within the Semantic Web [2]. The inability of machines to understand Web pages to a significant extend brought up several initiatives to overcome this weakness. Initially, the aim of the Semantic Web was to invisibly annotate Web pages with a set of meta attributes and categories in order to enable machines to interpret parts of the knowledge included in the text and to put it into some kind of context. This approach did not succeed since the annotation was rather complicated for
humans with no technical background and it therefore got stuck in the bootstrapping process. A current mark-up-based approach called Microformats [3, 4] follows a similar, but more atomic, idea of making the markup process easier and thereby allowing more users to participate and thus to overcome the cold-start problem. On the other end of the technology spectrum we find full-blown ontologies that describe domain knowledge with help of formal logic. This allows to infer new information from a set of facts, but is difficult to do with the right level of detail.

All these approaches have in common that they try to improve the machine-readability of web pages that are designed for humans. But the horizon or depth of machine-readable knowledge that can be added to a page is limited: only the page itself and particular elements on it can be marked-up by applying these approaches.

This limitation and the fact that nowadays there are already existing data sets containing lots of structured data about all kinds of information distributed over the world lead to the idea of creating a Web of Data: when these data sets are semantically described and interconnected, a machine can traverse through this web to gather semantic knowledge about arbitrary entities, independent of the information contained in the original Web page.

**State of the Art** A promising approach is the W3C SWEO Linking Open Data community project\(^1\) [5] (LOD). The project is uncoupled from the “Web for humans” and interconnects existing open data sets. Figure 2 illustrates currently connected data sets. The data sets contribute by granting access to their semantically linked knowledge and by linking to items of other data sets. This way, the project follows basic design principles of the World Wide Web [6], e.g. simplicity, tolerance, modular design and decentralization. The LOD project currently counts more than 2 billion\(^2\) RDF\(^3\) [7] triples.

The LOD data sets can be accessed using a Semantic Web browser or crawled by a spider of a semantic search engine, thereby empowering users to create very sophisticated queries. The search results are “pure knowledge” or matching entities from within the data web.

**Conclusion** With every fact and link added to the Web of Data, more and more general and specific knowledge is made accessible to machines. The Web of Data enables a new generation of search services. Through the semantic structuring of the data within the data sets and the interconnection of lots of different data sets, highly sophisticated queries become machine-processable and can be answered. Querying languages like SPARQL [8] and RQL [9] are already available.

---

\(^1\) <esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>, accessed Dec 2008

\(^2\) [www.w3.org/2008/Talks/WWW2008-W3CTrack-LOD.pdf](http://www.w3.org/2008/Talks/WWW2008-W3CTrack-LOD.pdf), accessed Dec 08

\(^3\) Resource Description Framework
Fig. 2. Data sets of the Linking Open Data project. The circles are data sets containing knowledge covering different domains. The thickness of the arrows illustrates a measure for the bi-directional connectedness. Image taken from LOD project.

2.2 Web of Services

**Definition 2 (Web of Services).** The Web of Services is a distributed web of semantically annotated services.

Fig. 3. Web of Services: Distributed, semantically annotated services are freely accessible to be e.g., discovered, invoked, orchestrated or chained. The arrows exemplary depict a possible service chain which might be discovered and invoked by an orchestrator.
**Problem Motivation** The Services sector has become the world’s biggest business sector forming 64% of the world-wide GDP\(^4\) [10]. This sector has a pressure to make their services easier and more widely accessible, as well as to adapt to ever faster changes in the market environment.

The Service Oriented Architecture (SOA) paradigm has become the predominant approach to (enterprise) software engineering, to streamline the IT infrastructure within an organization as well as to interact with external entities. Its principles [11] call for services that have their (formally described) interface decoupled from its functionality and described in an abstract fashion, to name just a few. While SOA can be implemented with a host of different technologies, web services have become the technology of choice.

However, it should be noted here that while services are all the rage, there is no clear definition as to what constitutes a service, neither on a technological level nor on a conceptual. In the former case, Web Services offer a quasi–standard, but the subset of standards that is agreed upon is neither powerful nor expressive enough to handle the possible applications of services. Also, the standards so far lack any semantic description, making its use within automated scenarios all but impossible. On a conceptual level, the situation is possibly worse as it is by no means clear what actually constitutes a service once we move out from the narrow definition of Web Service towards higher–level services.

Having said all that, today there are already all kinds of services with all levels of complexity on the Web and the number of them is expected to grow exponentially. The services follow different standards and a lot of them are proprietary, uni–directional and designed to be used by humans to mash–up something new from them. There are editorial catalogs, e.g. ProgrammableWeb\(^5\), designed for humans searching for a particular service. A lot of Web 2.0 services provide services to read existing or create new data exposing almost all of their functionality through their API (Application Programming Interface)\(^6\). There are even human–based services like Amazon’s Mechanical Turk\(^7\). As mentioned above, Web Services follow an agreed standard dealing with the service definition but currently lack of a semantic description. While there are a number of different approaches to adding a semantic description to Web Services, such as OWL-S [12], WSMO [13], or WDSL-S [14], none has so far managed to break out of its academic confinement.

Once services are annotated semantically, they get accessible by machines automating service discovery, execution, billing or revenue sharing, orchestration, replacement on failure based on experience (QoS\(^8\)) etc. These Web Services will be brought together in a Web of Services according to Web principles.

---

\(^4\) Gross Domestic Product
\(^5\) [www.programmableweb.com](http://www.programmableweb.com), accessed Dec 08
\(^6\) In 2007, Twitter counted ten times more traffic on their API than on their website ([www.readwritetalk.com/2007/09/05/biz-stone-co-founder-twitter](http://www.readwritetalk.com/2007/09/05/biz-stone-co-founder-twitter), accessed Dec 08)
\(^7\) [www.mturk.com](http://www.mturk.com), accessed Dec 08
\(^8\) Quality of Service
State of the Art  Many works deal with the topics Internet of Services and SOA in general, in research as well as industry, and we will focus here on the larger research projects. Closest to our idea is the SOA4All project\(^9\). It focuses [15] this goal through four corner stones. Firstly, Web principles, which we already noted in the context of the Web of Data, and web technology as the underlying infrastructure are used in the Web of Services. Secondly, they plan to implement user participation in terms of e.g. ranking of services. Thirdly, they want to facilitate Semantic Web technology to abstract from syntax to semantics to grant machines knowledge about the services. Last but not least it is planned to implement a context management to enable processing of user requirements when it comes to service contracting or orchestration.

The TripCom project\(^{10}\) concerns itself with the design and implementation of an architecture for application integration based on the combination of Semantic Web, Web Services, and tuple spaces, called the triple space service technology. There, services can persistently publish semantically annotated data in order to facilitate orchestration and choreography of services.

The SUPER project\(^{11}\) focuses on elevating business processes from the IT to the business level. To do so, appropriate ontologies are defined and services are semantically annotated to allow the context-aware automated integration of services in business processes.

The SHAPE project\(^{12}\) provides an unified approach to the definition of semantically enhanced SOA. The focus lies on the integration of model-driven approaches with semantics and SOA.

Somewhat orthogonal to the development of services is the large research area of multi agent systems [16]. There, similar to SOA — and actually predating it by at least ten years — distributed agents communicate and cooperate to achieve some goal. Where services are generally considered passive however, agents are autonomous (in the sense of having the ability to make own decisions) and proactive. Given some semantically described goal, an agent tries to bring about a situation where the goal holds true. He does this by interacting with other agents, cooperating with them to change the state of the world. In the context of the web of services, agents play a vital role in that they, at least in the realm of academia, already created a web of services, where machines (i.e. agents) autonomously searched for functionality and used different services based on their semantically described capabilities. The Agentcities [17] project tried to create a global, open, heterogeneous network of agent platforms and services to which any agent researcher could connect their agents. Services could automatically be offered and used.

Conclusion  The Web of Services will enable machines to work with a huge toolbox of functionalities. Services might answer queries (from humans or other

\(^9\) www.soa4all.eu, accessed Dec 08
\(^{10}\) http://www.tripcom.org, accessed Dec 08
\(^{11}\) http://www.ip-super.org/, accessed Dec 08
\(^{12}\) http://www.shape-project.eu, access Dec. 08
services), or create further knowledge which could also flow back to the Web of Data. Automated service orchestration and chaining will be an important tool to quicken innovation cycles.

2.3 Web of Identities

Definition 3 (Web of Identities). The Web of Identities is a distributed web about users: their personae, their social graphs and their assets. It provides privacy-preserving accessibility to user information.

Fig. 4. Web of Identities: A distributed web of identity hosters managing the identity, personae, social graphs and assets of their customers. The arrows exemplify identity linkage through the social graphs of the customers.

Problem Motivation The Web 2.0 [1] brought a paradigm shift in terms of user participation and contribution. The distinction between content consumers and content producers got blurred. The world-wide impact on user participation is immense: e.g., the world’s biggest Social Network Service (SNS) Facebook\(^\text{13}\) currently counts more than 130 million active users, is the most trafficked social media site and with 10 billion photos the biggest photo sharing application in the world\(^\text{14}\).

The major problem of these SNS or other web applications is conflict of interests in terms of property of data and privacy: the application providers put all efforts in user-base and content-base growth and currently usually own all the data that is added to their site. As the user–base and content–base directly influences the enterprise valuation, the companies are not interested in disclosing user or content details\(^\text{15}\). Currently, the users are not aware of these problems.

\(^{13}\) www.facebook.com


\(^{15}\) the companies’ behavior is called lock-in, walled garden or data silo
but as the amount of services that the users access growth, the downsides get obvious: the user has to re-enter his personal information, re-enter his preferences, re-enter his relationships to other users time and again. No site provides sophisticated import and export features and the user never controls, possesses or owns his own data.

To overcome this issues, we envision a solution which we call the Web of Identities. In it, a interconnected web of Identity Hosters (IH) take care of their customer’s data. An IH hosts the user’s data, e.g.

- identity (the user’s person)
- personae (a user must be able to differentiate between multiple personae, e.g. a private and a business persona)
  - profile data
  - social graph,
  - groups, lists
  - live-stream, instant messaging and other live data, messages
  - files, assets
  - reviews, comments, ratings, feedbacks
  - presence information, e.g. location, mood

and provides all needed management, privacy, access control, security, trust, AAA\textsuperscript{16} etc. features. Given the permission of the user, third party services (3PS), e.g. an SNS, can read and write needed fragments of this data. The user is in full control of his data being hosted and exposed. In this scenario, 3PS’s utilize the elements of the IH’s data-base they are allowed to access on behalf of the users who want to use this 3PS. The user profile, the social graph etc. are synchronized with the IH. The 3PS only implements the delta of data and functionality not being provided by the IHs. Each user of a particular 3PS can have identity information hosted by a different IH, while still interacting on one 3PS.

As the Web of Data and the Web of Services, the Web of Identities should follow basic Web principles.

It is essential that IHs emerging from both sides, bottom-up, e.g. a solution developed from scratch, and top-down, e.g. the players like Google or Facebook opening up more and more, can converge in an interoperable Web of Identities, i.e. IHs, speaking the same language. Like this, every user can decide which IH to trust and to choose as his personal identity hoster.

\textbf{State of the Art} There are bottom-up as well as top-down approaches, some are driven by commercial interests, some are non-profit. All of them are to a certain extend coherent with our vision.

First of all, we want to name the non-profit OpenID \cite{OpenID}, which initially focused on providing a distributed single-sign-on solution for authentication. The OpenID framework is designed with respect to the Web design principles and is completely open. The framework might be the identity and authentication fundament for a future Web of Services hoster implementation.

\textsuperscript{16} Authorization, Authentication, and Accounting
The big picture for all bottom-up approaches is drawn by Marc Canter with his Open Mesh\textsuperscript{17}. Canter outlines a vision of what building blocks are needed and how they could be put together in a common infrastructure.

The non-profit DataPortability\textsuperscript{18} group deals with the establishment of open standards and protocols for the exchange of data between applications and vendors. The protocols and standards are already widely agreed and now need to be further diffused and adopted. The development of open, non-proprietary specifications for web technologies is also the dedication of the non-profit Open Web Foundation\textsuperscript{19}.

A similar approach is taken by non-profit Identity Commons\textsuperscript{20}. They focus on the users’ identities and social graphs.

A distributed approach of providing SNS building block functionality is taken by the non-profit DiSo Project\textsuperscript{21} (Distributed Social Networks). The team implements a WordPress plugin that implements some of the standards supported by the DataPortability group.

A higher aggregated approach is the non-profit OpenSocial Foundation\textsuperscript{22}. It creates specification intending to spread social content and functionality across the Web following a centralistic approach.

The EU-funded project PICOS (Privacy and Identity Management for Community Services)\textsuperscript{23} focusses on creating a mobile meta-community solution. It covers some privacy and identity management challenges but, as a meta-community only, not at the scope of what we see with the Web of Identities.

Last but not least the big players are opening very slowly. On the one hand they do not want to expose data but on the other hand they still want to be useful for their users, so they open bit by bit. Features like Google’s Friend Connect\textsuperscript{24}, Facebook’s Connect\textsuperscript{25} and MySpace’s Data Availability\textsuperscript{26} do all aim at spreading a fragment of the platform’s features to outside of the platform itself. That serves the users’ needs of accessing locked-in data for a transitional time but does not solve the ownership issue. Microsoft’s Live Mesh allows it’s users to synchronize files across devices and platforms. Yahoo! opened up through the exposure of lots of services according to its Y!OS\textsuperscript{27} (Yahoo! Open Strategy).

\textsuperscript{17} blog.broadbandmechanics.com/2008/05/how-to-build-the-open-mesh, accessed Dec 08
\textsuperscript{18} www.dataportability.org, accessed Dec 08
\textsuperscript{19} www.openwebfoundation.org, accessed Dec 08
\textsuperscript{20} www.identitycommons.net, accessed Dec 08
\textsuperscript{21} www.diso-project.org, accessed Dec 08
\textsuperscript{22} www.opensocial.org, accessed Dec 08
\textsuperscript{23} www.picos-project.eu, accessed Dec 08
\textsuperscript{24} www.google.com/friendconnect, accessed Dec 08
\textsuperscript{25} developers.facebook.com/connect.php, accessed Dec 08
\textsuperscript{26} developer.myspace.com/community/myspace/dataAvailability.aspx, accessed Dec 08
\textsuperscript{27} developer.yahoo.com/yos, accessed Dec 08
It is important to note here, that the big players Google, Yahoo, Facebook, Microsoft and MySpace are all for themselves on the way of becoming one of the envisioned IHs.

**Conclusion** For the Web of Identities there is no big, well focused project like LOD for the Web of Data and SOA4All for the Web of Services, yet. We see the attention and need for this solution slowly rising and all the named activities converging in the Web of Identities. We are sure the directions and corrections will be driven by user needs and the market.

Research has to be done in the areas of empowering the user to take control of his data. Features like reach control, revokable access rights and the management of which third party service can read or write what fragment of which persona’s data are necessary but very hard to translate to an intuitive user interface and –experience.

If this vision comes true, we will see a user centric, user friendly, privacy preserving and meaningful tool. Users can explicitly grant online marketers access rights to attention data or purchasing history data to empower them to target meaningful ads that may take into account what direct friends recommend.

For machines, the Web of Identities is a very important infrastructure for looking up user-related private, volatile personal and contextual data.

3 Interplay of the Webs

With the Trilogy of Webs as a backbone, the Internet as a tool will change because through interconnected knowledge and toolsets, machines are catapulted to a new ability level. New services will emerge basing on the fundament of the Webs. From the HCI\textsuperscript{28} side, the way of how we use existing services will change dramatically.

The distinction between data items stored and retrieved and the use of services will continue to blur until requests will freely traverse the webs, retrieving items of data, feeding chains of services that use personal information from the stored identities.

The following scenarios give an impression of how the webs interlink.

3.1 Scenario: Search

A user, let us call him Peter, visits a search engine system that is based on the Webs. Peter queries *Recommend books about Berlin for my mother for Christmas*. The NLP\textsuperscript{29} component of the system analyzes his query. The query is split up to a chain of tasks, the system now starts to process:

From the Web of Data, the system gathers general knowledge about the terms *mother, Berlin, Christmas*. The system now asks the Web of Data about

\textsuperscript{28} Human Computer Interaction
\textsuperscript{29} Natural Language Processing
all books covering Berlin or authors born or living in Berlin. Given Peter’s permission, his IH is called to return his mother’s identity URI\(^{30}\) from his social graph. Peter’s IH searched all of Peter’s personae for his mother and found her in his private persona’s social graph. The mother’s IH is called to access her interest information limited to the wider topic fields books and Berlin. The mother’s personae’s social graphs are searched for her backlink to Peter. As the private personae is found, the private personae’s information are selected for access limitation. From the private persona, the mother’s IH returns a set of information the mother explicitly granted access to. The set contains general interests, some purchases, reviews, comments, ratings and some attention data. URIs of the mother’s closest friends are also returned. The system continues by querying the mother’s closest friends’ IHs if one of them liked or recommends books about Berlin since friends’ recommendations are the most valuable. The system identifies the term recommend as a service request and searches the Web of Services for a recommendation service that can handle books and personal interests and recommendations as filtering and ranking criteria. The initial set of books the system retrieved from the Web of Data and the information collected from the Web of Identities are now sent to a filtering and ranking service. As the term Christmas is recognized as contextual term for the task, the system now searches the Web of Services for e-commerce services offering books. The filtered and ranked list of books is sent there to offer prices and delivery dates before Dec 24th.

The list of books augmented by prices and dates is presented to Peter. The system tracks Peter’s feedback for the book recommendations and assigns it to its QoS ratings for the services it invoked.

\(^{30}\) Uniform Resource Identifier
3.2 Scenario: Mass Customization

Peter just graduated from university. He knows that he needs an insurance package but he has no idea what exactly it should consist of. He heard of an intelligent insurance brokerage system which he now visits with his browser. He logs into the system via his IH. From the Web of Identities and with Peter’s permission, the system initiates a profile lookup at Peter’s IH to gather information needed for the configuration of the components of the insurance package. It queries for information like private address, marriage status, age, gender. Since it cannot find Peter’s current income, it prompts Peter directly. From the Web of Data, the system now queries for Peter’s neighborhood’s crime statistics for later risk estimates. The system now looks up all insurance services it can find in the Web of Services. It configures the services with the knowledge gathered, selects the best offers and combines them in a personalized insurance package for Peter. The package consists of a handful of insurances from different insurers around the world. Peter signs the contracts through the broker and logs out with the satisfaction that he now is optimally and neither under- nor overinsured.

3.3 Believe, Desire, Intention

If you will, the Webs can be compared to the BDI (believe, desire, intension) model of Rao and Georgeff [19] that describes a formal model of the mental state of agents. They describe the mental state of a single agent with help of a (incomplete) model of the world (i.e. beliefs), a set of plans (i.e. possible courses of actions, or intensions) and a set of goals — desired states of the world. On a global scale, the Web of Data represents the believe state of the world, while the services and their composition provide possible courses of actions, and identities contain goals and desired states. In this reading, the Webs of data, services, and identities move indeed the world a bit closer to machine understanding.

4 Conclusion

In this paper we have outlined the concepts of the Web of Data and the Web of Services, and introduced the Web of Identities. We have demonstrated how, in parallel to the Web for humans, these interplaying Webs will provide a new level of machine understanding and interoperability which one could see as common sense for machines.

We want to note that it is indispensable that all Webs ensure security, privacy and trust, internally as well as in their interaction. The notions as described here however allow to include any mechanisms that support these, just as the world wide web provides the basis for secure transactions without prescribing technologies.

Our vision of three interlinked yet clearly definable areas within the future internet allow for focussed research and development in either each of the webs or in their interactions, some of which we hinted at in the scenarios. However, other interactions are of course possible. We believe that research areas and business cases can and will arise from the web of machines as described in this paper.
References