Is data sharing the privilege of a few?
Bringing Linked Data to those without the Web

Christophe Guéret, Stefan Schlobach, Victor de Boer, Anna Bon, and
Hans Akkermans

Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081HV Amsterdam
{c.d.m.gueret,k.s.schlobach,v.de.boer,a.bon}@vu.nl,
Hans.Akkermans@akmc.nl

Abstract. Linked Data has been designed on the basis of the Web of
Documents and pushed by individuals with privileged access to Web
technologies. As a result, Web hosted applications and data servers have
become the *de-facto* standard for setting up data-sharing platforms. But
what about the estimated 4.5 billion underprivileged without access to
the Web but with a need and the willingness to share data? This paper
calls for taking in account *everybody* in future design choices for Linked
Data. We claim that "Web-free" publication is possible, and beneficial,
for Linked Open Data everywhere - even without access to the Web.

1 Introduction

The World Wide Web has had an enormous impact on the economic and democratic situation around the globe. However, as it is recognised by its inventor, Tim Berners-Lee, and others: not everybody can benefit. While Web access in rich countries has reached near universal numbers, an estimated 4.5 billion underprivileged people in poorer areas of the world remain excluded. The recently formed Web Foundation[1] recognises and tries to bridge this huge 'digital divide'.

While the Web of Documents has been around for 20 years, the Web of Data is currently still being engineered. We, engineers of Linked Data, have the unique opportunity - and moral obligation - to consider the underprivileged majority as we make design choices. Rather than developing solutions that only benefit the privileged few and worry about the rest of the world later, we should empower every citizen on the planet through our Linked Data solutions.

The main obstacle to bring Linked Data to the poorer rural areas of the world is the lack of infrastructure. The current strategy has been to build the Web of Data on top of the infrastructure of the Web of Documents. Web hosted applications and data servers have become the *de-facto* standard for data-sharing platforms - and the Cloud computing trend will contribute to strengthen this.

In order to turn the Linked Data into a phenomenon for the masses Semantic Web technologies have to be deployed locally. As an alternative to the centralised publication of data on the Web, we propose a new model which can revolutionise

---

[1] www.webfoundation.org

the publication of Linked Data. This model is based on micro-servers enabling local information sharing while preserving global connectivity.

2 Data sharing for individuals without Web access

ICT has the potential to play a major role in social development and the empowerment of people. Yet, 4.5 billion people, mainly in developing countries, do not have access to the Web. These individuals do however produce and generate data and knowledge they are willing to share, even without having access to the infrastructure of the Web. They can benefit from Linked Data *right now*!

Example 1: Rural communities in the Sahel In rural areas in *e.g.* Burkina Faso and Mali, communication is voice-based, since many people cannot read or write. Mobile phone and community radio are the main channels for information and communication. To provide a voice-platform for local knowledge sharing, mobile voice services are currently being developed and deployed locally, as part of the EU-funded VOICES project[2]. These services are developed in close dialogue with people from rural communities, local radio stations, and local ICT entrepreneurs, as to make them useful in the local context. The voice-services are hosted locally, on "micro-servers" that can be accessed by simple first-generation mobile phones, through inexpensive local phone-lines [1].

One such voice-service allows, *e.g.*, to issue a voice message that is resent to several phones, like a sort of "voice-Twitter". People can react on it and access or leave a voice message. People can receive the voice message translated into their own language, based on their user-profile. The voice-messages are stored in a repository, hosted on the local micro-server. This information might be interesting to people in other regions as well. By opening the repositories of voice messages using Linked Data principles to similar services in other regions, the information can be shared over a wider area and serve more people. As RDF is particularly suited to abstract over multiple languages and facilitate data integration, Linked Data would enable this knowledge sharing capabilities.

Example 2: Children collaborating in knowledge creation "One Laptop Per Child" (OLPC) is a non-profit organisation whose goal is to empower the world's poorest children through education by providing them with a rugged, low-cost, low-power and connected laptop suitable for usage in developing countries: the XO. This device connects them to each other and to the World via the Internet. The learning platform "Sugar" powering the XO promotes collaborative learning through activities that encourage critical thinking. It has been designed from the ground up especially for children, is available in 25 languages and is being used daily by almost two-million children in more than forty countries.

Collaboration is a key aspect of Sugar which translates into extended sharing capabilities between activities. The drawing activity, for instance, allows for many children to collaborate on the same drawing at the same time. The same

---

[2] http://www.mvoices.eu

applies for the text editor, the discussion rooms and others. This data sharing is still however activities-centric. Moreover, the current data model used in Sugar doesn't easily cope with multi-lingual information. Not as easily as RDF does.

The community project SemanticXO [2] which started in October 2010 is about deploying semantic technologies in Sugar to create new data sharing opportunities across applications. Although school server are used in some schools, we can not consider having a triple store per school dedicated to storing the triples generated by the activities. Besides, the XO are meant to be carried home by the children who owns the device and the connectivity to a school server could not be ensured. The solution taken in SemanticXO is to setup a dedicated triple store directly on the XO.

3 Linked Data with Web-free micro-servers

Bringing Linked Data to the masses yields both technological and societal issues. As Linked Data engineers we hereafter focus on the first aspect but we acknowledge that micro-servers won't solve everything and suggest to phrase this duality as a call for multidisciplinary research.

The design principles of Linked Data allows for a given data set to be seen as a collection of triples. Classically, a single machine - eventually backed up by a cluster - is used to store all the triples and is accessed through the Web. But this set of triples can also be scattered over, and edited by, several machines acting individually but collectively reproducing the behaviour of a centralised, Web hosted, service. By analogy with the concept of "microgrid" for energy production[3], we refer to these machines as "micro-servers" and present them as an alternative to centralised hosting services.

Micro-server architecture Micro-servers are machines with limited capacity aimed at storing and serving locally relevant Linked Data. They are connected devices that can exchange data with each other and eventually connect to the Internet whenever such connection is available. Considering the rapid progress of mobile technologies, we envision concrete implementation of micro-servers for Linked Data as phones or tablets applications and/or operating system extensions in the near future. We note that these low-powered devices with passive cooling and no moving parts are particularly suited to usage in African countries where dusty environment and lack of electricity can be an issue.

A first micro-server is under active development under the context of SemanticXO. This micro-server is based on the software stack of the XO and is implemented in one of these low-end laptop designed for usage by kids in developing countries. Our test device is an XO of the first generation with a 433Mhz x86 CPU, 256MB of RAM and 1GB of flash memory [4]. First tests showed that this device can be used to store 50k triples while answering basic SPARQL queries[2]. The performances later observed showed satisfactory performances for practical use as no significant latency was observed in the developed activities.

---

[3] http://en.wikipedia.org/wiki/Distributed_generation#Microgrid
[4] http://laptop.org/en/laptop/hardware/specs.shtml

Challenges As one would expect, there are challenges to overcome when aiming at using Linked Data without the Web and using a swarm of micro-servers instead of central servers. The proposed de-centralised publication scheme breaks with the traditional centralised control of data and thus poses new problems with respect to security, consistency, reliability and de-referencability of the data:

– **Security** We must ensure that micro-servers are legitimate data providers of the data they serve. Without a central authority, any peer could potentially declare itself a data provider for any data set.
– **Consistency** Every micro-server produces data and thus may come to minting URIs or re-using others. There must be a control mechanism ensuring that URIs are re-used across micro-servers when applicable and that no collision happens for minted URIs.
– **Reliability** A solar-powered micro-server deployed in a rural area can not have the same guaranteed up-time as a grid-powered server deployed in a data center. There is thus a need for data duplication amongst the swarm of micro-servers to increase data availability and provide reliable persistence.
– **De-referencability** Considering Linked Data without the Web also removes the presence of DNS, used to resolve mappings from host names to IPs. An alternative will have to be found to provide HTTP de-referencability without relying on the usage of paid level domain names (PLDs) which implies an access to the DNS architecture.

4 Looking at the future...

Linked Data represents a significant progress with a huge potential impact. But to *really* make a difference everyone must gain access to it. 10 years of developments and design choices focused around Web systems and centrally hosted triple stores have kept Linked Data out of reach out 4.5B of the world's citizens. These individuals are also potential data consumers and producers (the so called "prosumers") that could benefit from the advances Linked Data represents in terms of Knowledge representation and reasoning.

We do have the opportunity to adjust our design principles, best practices and guide-lines to ensure Linked Data will not become the privilege of less than a third of the world's population. Let's seize the opportunity while Linked Data is still in its infancy to make sure that everyone can benefit from it.

References

1. Akkermans, H., Grewal, A., Bon, A., Tuyp, W., Allen, M., Gyan, N.B.: W4ra-voices field report. Tech. rep., Web Alliance for Regreening Africa (2011), http://www.mvoices.eu/2011/03/25_Voices-W4RA_Public_Report.pdf
2. Guéret, C., Schlobach, S.: SemanticXO : connecting the XO with the World's largest information network. In: Proceedings of the First International Conference on eTechnologies and Networks for Development ICeND2011. "Communications in Computer and Information Science", Springer LNCS (2011), http://bit.ly/p9jrDg