Decentralised Open Data for world citizens

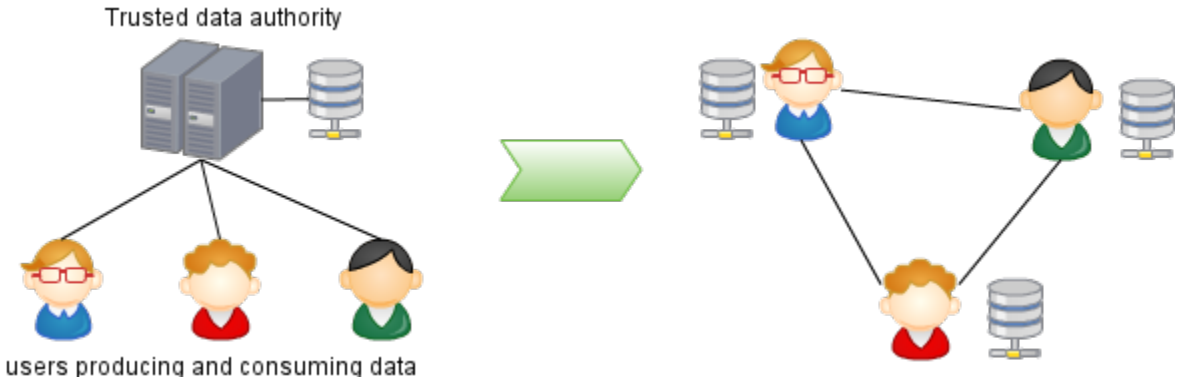
Christophe Guéret, Victor de Boer, Anna Bon (Vrije Universiteit Amsterdam, The Netherlands)

Abstract

Open Data production and consumption is currently considered essentially through access to centralised data servers. This short position paper highlights the shortcoming of this approach for world citizens having a limited access to network infrastructures and cites the work done at the VU University Amsterdam (VUA) to tackle this issue.

Towards decentralised publication and consumption of open data

In a lot of contexts, for example in the specific case of developing countries, access to network infrastructure is limited. Citizens may not have the reliable access to the Internet nor a stable electricity grid needed to operate the applications that use of the open data they want to consume. To ensure access for all citizens to Open Data published by the government as well as individuals, access patterns have to be decentralised, switching from a central read/write point to several distributed ones:



At VUA several initiatives were taken to realise decentralised data publication and consumption to empower people living in developing countries. Among the different projects, the following are of particular relevance:

- RadioMarché: a voice-accessible market information system in rural Africa
- Foroba Blon, decentralised open data for community radios in rural Africa
- SemanticXO: implementation of a data management layer for the XO distributed by One Laptop per Child (OLPC). This data management stack allows for the publication and sharing of data directly from the XO, without having to rely on centralised storage.

In the following sections we give describe examples of decentralised data publication in a development context.

RadioMarché: a voice-accessible market information system in rural Africa

Market Information Systems are of great value for subsistence farmers in remote rural areas, because they can match their small scale product offerings with the demand from potential buyers [1].

RadioMarché is a Market Information System adapted for rural conditions in the African Sahel. The MIS is specifically dedicated to promoting sustainable use of forest resources and develop small businesses based on so-called "non-timber forest products" such as shea nuts, shea butter and honey. The system is currently used in the Tominian region of Mali to distribute up-to-date market information via community radio in the area. Up-to-date market information such as product offerings (including product type, quality, quantity, location and logistical issues) is accessible through a Web interface or through a voice interface for second-generation mobile phones.

To maximize the reusability across multiple decentralized instances of RadioMarché, but also across different application domains and regions and to allow for automatic machine processing, we adopt Linked Open Data standards to share the market data [2]. We are currently enriching this data with links to other open data sources, identifying interesting use cases for re-use and developing prototype services using this linked market data [3]. Although the data is currently hosted on a single server in Amsterdam, we are investigating opportunities to host this data on low-powered hardware in the region itself.

Foroba Blon: decentralised open data and African community radio

In rural regions of West-Africa community radio plays an important role as information provider. In the absence of reliable electricity grids, stable internet connections and computers, most people rely on the radio as their main source of daily information. Community radios are often small enterprises, broadcasting within a radius of 100 km on average. Community radio stations create their own programs and broadcast local and regional news, music, informative programs, round-table programs and paid announcements, mainly in regional languages. The listeners of the radios are poor farmers, who do not have access to the internet. Their data consume is based on radio as the interface. Data production is done through mobile phone, by phoning to the radio and leaving a spoken message for broadcasting.
Currently, a decentralized radio platform is being developed for African community radio [4]. It will contain a data store of local recordings, i.e. audio files and their associated meta-data, produced by the listeners, village reporters and local journalists by mobile phone. The radio platform allows both web and phone (voice) access. This radio platform is currently being developed for three community radio stations in Mali.

The radio platform could in theory be physically hosted anywhere in the world, on any web server, connected to the Internet, on a central server, hosted by one of the trusted data authorities. However,

in the case of rural regions in Africa, this is not possible. Firstly, the radio platform has to be accessible using an inexpensive local phone number, so it must be connected to a local African phone network. Secondly, the web service accessed over the Internet must also be accessible locally. Since the internet connectivity in this part of the world is usually of low bandwidth and of high latency, web services hosted in datacenters in the US or Europe, are too slow for proper deployment. Therefore the system has to be hosted in the vicinity of the end-users, in the same country or in the region. In the absence of good and reliable datacenters or hosting providers in West-African countries the radios prefer to deploy the radio platform decentralized, in this case having a local server at their own premises.

To open the data of the radio platform, Linked Data principles will be applied for publishing and connecting data on the Web, while adhering to its architecture and standards so that people can discover more things, especially across regions and similar community radio platforms. When the radio platform that is currently deployed in Mali proves to be a success, other instances will be installed at local radio stations in other rural regions of Africa, across borders, in neighboring countries, Burkina Faso, Ghana, Senegal where conditions with regards to illiteracy, local languages, mobile telephony and community radio are similar. The decentralized different instances of the radio platform will be linked together, so that they can serve the local communities and be adjusted to the local use.

Semantic XO: Children collaborating in knowledge creation.

One Laptop Per Child (OLPC) is an initiative for the empowerment of 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. The learning platform Sugar powering the XO promotes collaborative learning through activities that encourage critical thinking. XO has been especially designed for children and is available in twenty-five languages. Currently, almost two-million children in more than forty countries are using the XO at school, every day. Collaboration is a key aspect of Sugar which translates into extended sharing capabilities between activities. The drawing activity, for instance, allows children to collaborate on the same drawing simultaneously. The same applies for the text editor, and other applications. The original data sharing of Sugar is still activity-centric. The community project SemanticXO, started in October 2010, aims to deploy semantic technologies in Sugar, to allow data sharing across different applications [5]. Since many schools in developing countries do not have school servers, a local network or a connection to the internet, the data stores are kept on each individual XO. Besides, the XOs are meant to be carried home by the children. Therefore, SemanticXO data management stack allows for the publication and sharing of data directly from the XO, without having to rely on centralised storage.

Discussion

In this paper we argue in favor of more decentralization of Open Data, for the empowerment of world citizens. In the past few years, many efforts are made to connect people in developing countries to the global information society using ICTs. Despite a multitude of efforts, only a few were dedicated to the use of Linked Open data in a development context. In the three cases described in this paper, the lack of

a proper infrastructure was addressed by deploying decentralized solutions.

In the case of the rural market information system, using a centralised server, would not be appropriate because of lack of local infrastructure. Therefore, the data production, consumption and storage is decentralized towards the community level. Similarly, in the community radio project, a decentralized solution seemed much more favorable. In the case of Semantic XO, the datastore was even decentralized to the level of the classroom and the individual school children, to make the data sharing possible, even in a server-less school environment.

Still, in order to leverage the full potential of decentralized Open Data, several challenges need to be addressed, especially those related to trust and coherence of the data. In a model where everyone could pretend being the source of the data published, means of provenance tracking and data integrity are of particular importance. Furthermore, the possibility to cache and reserve data that was previously downloaded adds to this authenticity problem and data freshness issue.

Yet, there are many advantages of decentralisation. Besides enabling access to data, the proposed decentralised peer-to-peer model will significantly diminish the load from individual network resources. This model will give more power to the users, as they retain more control over the data they provide. It can be noted that a hybridization of the model is possible with users relying both on external services (such as e.g. Amazon Cloud or Google Drive) and on decentralised servers, to store and serve their data.

References

1. FAO: Financing agriculture and rural development in africa: Issues, constraints and perspectives. In: Twenty-third Regional Conference for Africa, Johannesburg, South Africa, 1-5 March 2004.
2. De Boer, V. De Leenheer, P. Bon, A., Gyan, N.B. Van Aart, C. Guéret, C. Tuyp, W., Boyera, S. Allen, M. Akkermans H. RadioMarché: Distributed Voice- andWeb-interfaced Market Information Systems under Rural Conditions. Accepted for publication in Proceedings of 24th International Conference on Advanced Information Systems Engineering, CAiSE'2012, Gdansk, Poland, 25 – 29 June 2012.
3. De Boer, V. Gyan, N.B. De Leenheer, P. Bon, A. Van Aart, C. Guéret, C. Tuyp, W. Boyera, S. Allen, M. and Akkermans, H. Bringing the Web of Data to Developing Countries: Linked Market Data in the Sahel. Accepted for a poster presentation at ESWC 2012, Heraklion, Greece, 27 - 31 May 2012.
4. Bon, A. De Boer, V. De Leenheer, P. Van Aart, C. Gyan, N.B. Froumentin, M. Boyera, S. Allen, M. Akkermans, H. The Web of Radios - Introducing African Community Radio as a new interface to the Web of Data. Proceedings of the 1st International Workshop on Downscaling the Semantic Web, Heraklion, Greece, May 28, 2012.
5. 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).