Voice-based Access to Linked Market Data in the Sahel

Victor de Boer, Nana Baah Gyan, Anna Bon, Pieter de Leenheer, Chris van Aart, Hans Akkermans

Dept. of Computer Science, the Network Institute, VU University, Amsterdam, The Netherlands
{v.de.boer, n.b.gyan, a.bon, pieter.de.leenheer, c.j.van.aart, j.m.akkermans}@vu.nl

Abstract. In this paper, we present our ongoing efforts to bring the Web of Data to rural communities in the Sahel region. These efforts center around RadioMarché, a market information system (MIS) which can be accessed using first-generation mobile phones. We argue that linking the locally produced and consumed data to (external) Linked Data sources will increase its value. We describe how RadioMarché data is available as Linked Open Data and present a prototype demonstrator with voice-based access to this linked market data. Through this interface, the Linked Data can be accessed using first generation mobile phones. As such, these are first steps towards opening the Web of Data to local users that do not have appropriate hardware to produce and consume Linked Data. We present a number of use cases as well as the current deployment state. We also discuss our current efforts to leverage the creation of Linked Data in development regions and build applications on this Linked Data.

1 Introduction

Development and use of the Web of Data has until now mainly focused on developed countries, as was the case with the Web of Documents before it. 4.5 billion people - mainly in developing countries- currently can not access the World Wide Web. The reasons for this include infrastructural ones such as a lack of high bandwidth Internet connections and reliable power supplies as well as socio-economic issues such as the high cost of buying Personal Computers, language mismatches and lack of reading and writing abilities. For our case study in Mali, only 1.8% of the population has Internet access[1], only 10% has access to the electricity network[2], and only 26.2% is literate[3]. Currently, a number of efforts are being undertaken to bridge this so-called 'digital divide' in the World Wide Web, including the recent forming of the Web Foundation. As was argued in [1], while the Web of Documents has been around for 20 years, as engineers of the much newer Web of Data, we have the opportunity to not let the "digital Linked Data divide" grow too large. To avoid a seemingly unbridgable gap, we should consider the underprivileged majority as we design Linked Data architecture, describe use cases and provide access to that Linked Data. In this paper, we describe our ongoing

[1] http://www.internetworldstats.com/ Internet World Statistics, Miniwatts Marketing Group.
[2] http://www.developingrenewables.org/energyrecipes/reports/genericData/Africa/061129%20RECIPES%20country%20info%20Mali.pdf
[3] http://www.indexmundi.com/facts/indicators/SE.ADT.LITR.ZS Index Mundi 2011.

investigations the implementation of Linked Data-backed solutions for the rural Sahel regions.

1.1 RadioMarché

Our efforts center around a Market Information System, RadioMarché [2], a web-based market information system being developed within the VOICES project [4] aimed at stimulating agricultural trade in the Sahel region. The RadioMarché system augments an already running Market Information System (MIS), that was introduced by our partner NGO, Sahel Eco[5], in the Tominian Area in Mali.

Within RadioMarché local market information about Non-Timber Forest Products (NTFPs) such as honey, tamarind and shea nuts is stored. A local instance of RadioMarché has a data store with rudimentary market information such as product offerings (including product type, quality, quantity, location and logistical issues) and contact details from sellers and buyers. This information is sent to community radios for radio broadcast and is made available for individual potential buyers and sellers. To overcome interfacing and infrastructural issues, RadioMarché has a voice interface which can be accessed through the normal telephone network using first-generation mobile phones.

A first version of RadioMarché has been deployed in November 2011 in the Tominian region. In this paper we describe a prototype demonstrator developed in parallel which exposes the market data from this prototype instance of RadioMarché using Linked Data approaches, so that new opportunities for product and service innovation in agriculture and other domains can be unleashed. The prototype demonstrator also features rudimentary voice-access to the Linked Data.

1.2 Why Linked Data?

We believe that Linked Data as a paradigm is very much suitable for knowledge sharing in developing countries. Linked Data approaches provide a particularly light-weight way to share, re-use and integrate various data sets using Web standards such as URIs and RDF. It does not require the definition of a specific database schema for a dataset [3]. We assume that the majority of the use of the locally produced data will also be consumed locally. Although the specifics of the locally produced data will differ from use case to use case and from region to region, Linked Data provides us with a standard way of integrating the common elements of the data. Also, because we do not impose a single overarching schema on the data, data reuse for new services is easier, both within a region and across regions. The aggregated data can be used by NGO's to assess running programs and increase their own transparency and accountability. We will provide examples in Section 4.

An additional advantage is that Linked Data is well-suited to deal with multiple languages as its core concepts are resources rather than textual terms. Where the Web of Documents, by design, is language-specified, Linked Data is designed to be "language

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

[5] http://www.saheleco.net

agnostic", which suits our purpose of multilingual and voice-based access well. A single resource, identified by a URI (ie. http://example.org/shea_nuts) can have multiple labels (eg. Shea Nuts@en and Amande de Karite@fr). Other than textual labels, for our voice-services we add audio to the resources with language-specific voice snippets, also identified through URIs.

2 Related Work

Agarwal et al. from IBM Research India, developed a system to enable authorship of voice content for 2G phone in a Web space, they named the WWTW (World-Wide Telecom Web). The whole system creates a closed web space, within the phone network. Linking from one voice site to the other is done through a protocol HSTP, created by IBM. Especially the lack of open search possibility constrains its growth [4].

Several automated market information systems have been developed and built to support farmers and agricultural trade in developing countries. One of the well-known market information systems is ESOKO [5], an online market system, developed and built in Ghana. ESOKO enables sellers and buyers to exchange market information. Google started a project in Uganda in 2009, partnering with MTN and Grameen Foundation to develop mobile applications that serve the needs of poor and other vulnerable individuals and communities, most of whom have limited access to information and communications technology [6]. This system is based on SMS but does not allow voice access.

The Web Foundation has started the Open (Government) Data to "Conduct country level actions and global actions to increase the impact and benefits of Open Data worldwide" [7]. This effort focuses on opening government data in developing countries such as Ghana. Our data is initially designed to be produced and consumer by the regional farmers themselves. Linking our regional data to the (Linked) Open government data could increase the value of both datasets. A related project on Linked Data for developing countries is described by Guéret et al. [8]. The SemanticXO is a system that connects rugged, low-power, low-cost robust small laptops for empowerment of poor communities in developing countries.

3 The RadioMarché linked market data Demonstrator

3.1 The Linked Market Data

For our experimental demonstrator, we transcribed a copy of the up-to-date market information from the RadioMarché prototype deployed in the Tominian region to RDF triples and stored in a ClioPatria triple store [9]. The transcription process is done using XMLRDF rewrite rules [6], the conversion can be run when the RadioMarché database is updated to ensure the database of the deployed version and the linked data store of our prototype are in sync.

[6] http://cliopatria.swi-prolog.org/packs/xmlrdf

Currently, we use PURLs for the resource URIs. The temporary namespace chosen is http://purl.org/collections/w4ra/radiomarche/. An HTTP request to these PURL URIs is redirected to the ClioPatria server, running at http://semanticweb.cs.vu.nl/radiomarche/.

Through ClioPatria's Linked Data package, the RDF data is accessible as Linked Open Data. The result of an HTTP request for a resource is either a human-readable web page [7] or the raw RDF triples describing the resource (in the case of a browser request or an RDF request respectively). A SPARQL 1.0 endpoint is also provided at http://semanticweb.cs.vu.nl/radiomarche/sparql/.

As of February 2012, 31 market offerings are in the triple store. These market offerings have been done by 15 different farmers, living in 13 different villages spread across 6 regional "zones". The market offerings contain the quality, quantity and type of the product the price and contact information. In total, the market data consists of 721 triples.

In the current version of the demonstrator, the FOAF ontology has been used to describe persons. Additionally explicit links from the dataset to external data sources were made manually. These include links from zones and villages to GeoNames concepts, DBpedia geographical resources and DBPedia product descriptions.

3.2 Voice-based access to Linked Data

The linked market data can be browsed through the web using the above mentioned URLs. However, as stated, our goal is to provide a voice-based interface that allows non-intrusive market information access for all users having a first-generation mobile phone.

We have implemented a rudimentary version of a voice-based interface to the linked market data as described in the previous section. The voice service is built using VoiceXML [10], the industry standard for developing voice applications. Although in a deployment version we cannot assume that text-to-speech (TTS) libraries are available for the local languages, we here only implement English-language access to the data, using English TTS. Within the VOICES project, TTSs are currently being developed for local dialects of French as well as local languages such as Bambara and Bomu.

The prototype voice application is running on the Voxeo Evolution platform [8]. The platform includes a voice browser, which is able to interpret VoiceXML documents, includes (English) TTS and provides a number of ways to access the Voice application. These include the Skype VoIP number +990009369996162208 and the local (Dutch) phone number +31208080855.

When any of these numbers is called, the voice application accesses a VoiceXML document hosted on a remote server. This document contains the dialogue structure for the application. In the current demonstrator, the caller is presented with three options, to browse the data by product or region, or to listen to the latest offering. The caller presses the code on his or her keypad (this is Dual Tone Multi-Frequency or DTMF). The voice application interprets the choice and forwards the caller to a new voice menu.

[7] For example http://purl.org/collections/w4ra/radiomarche/village_Samoukuy/ shows all information about the Samoukuy village

[8] http://evolution.voxeo.com

For products, the caller must select the type of product (“press 1 for Tamarind”, “press 2 for Honey”, etc.), for regions the caller is presented with a list of regions to choose from. Based on the choice the application then accesses a PHP document on the remote server, the choice is copied as a HTTP GET variable.

Based on the choice, a SPARQL query is constructed. This SPARQL query is then passed to the RadioMarché Linked Data server, which returns the appropriate results. For a product query, all (recent) offerings about that product are returned. The SPARQL result is then transformed into VoiceXML and articulated to the caller.

The prototype demonstrator and the ways of accessing it are shown in Figure 1.

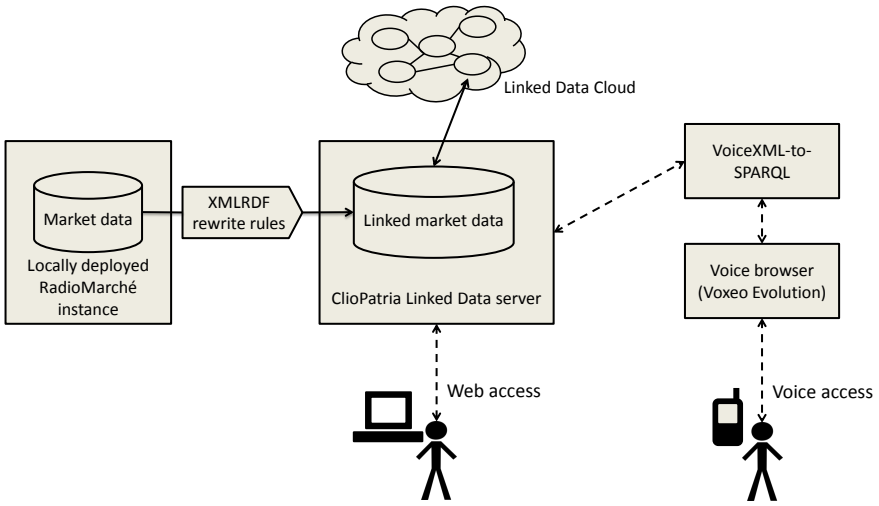

Fig. 1. Schematic representation of the linked market data prototype demonstrator.

Of course, the current method of accessing the data is only one of many possible actions. The caller can be presented with advanced filtering options (“enter the maximum price for offerings of product X”, “enter a date range for product offerings”) or combinations of data queries. However, because of the slow and linear nature of voice interfaces -when compared to visual UIs- options have to be limited more than with visual interfaces. This means that in our research we will identify useful services on this data and provide Voice-to-SPARQL mappings for these services.

4 Current Work

Voice-access to the linked market data as described above is still very much in an early prototype state. We are currently working on multiple projects to a) expand number of interlinked datasets produced and consumed in the region and b) investigate use cases

that use this Linked Data and build services and applications for those use cases. In this section, we describe the current status of these efforts.

4.1 Other Linked Data sets

The following is a list of Linked Data sets currently being realized. Each of these will be related to the linked market data as well as to external sources.

Meeting Scheduler Within the VOICES project a second use case is to develop a voice-accessible meeting scheduling system. The goal of this system is to provide local NGOs with a more effective way to transfer agricultural knowledge about non-timber forest products to their farmer community. The services developed in this case study provide voice access to personal and scheduling information. By integrating this information with the market information from RadioMarché, personal profiles can be enriched with information about the type of products that specific farmers have been producing within a given period. Here a new scheduling and notification service can re-use the market information within a region.

Citizen Journalism Data A second use case that is currently under development by the same team is a voice-based journalism platform, IPI innovation fund, which allows both professional and citizen journalists to send voice-recorded news items to local community radios. The target region for this use case consists of agricultural communities and there is a large possibility for re-use of both technical infrastructure as well as data. To do this, the re-usable resources (e.g. person data, geographical or product information) in the market information data are linked to the relevant resources in the target data set using Linked Data standard relations.

Pluvial Data We are also developing a crowdsourcing platform to transform photocopied data about rainfall in the Bankas area in Mali to Linked Open Data. This platform targets the 'diaspora', e.g. people originally from the region that have since moved to developed countries, where they might have better access to web browsers. The pluvial Linked Data acquired in this way will be linked to the aforementioned data. This can be exploited by our partner NGO as well as other NGOs to analyze for example patterns between rainfall and market offerings.

IDS Data The Institute for Development Studies recently published an API exposing more than 30.000 publications about development research [9]. As part of a recent agreement, we will develop a wrapper around the IDS API to expose its content as high quality Linked Data, enriching it with connections to other Linked Data datasets. These include both general datasets such as DBPedia or GeoNames as well as datasets with information from developing countries that are currently being realized. We will also develop a client application showing the advantages of this publication, exploiting the integration of the IDS data with other Linked Data sets in an information mashup.

[9] http://api.ids.ac.uk

Links to External datasets At the same time, local and national governments as well as NGOs can exploit the linked market data for analytic purposes, monitoring the trade in NTFPs within and across regions. By linking the market information to existing agricultural vocabularies such as FAO's Agrovoc thesaurus[10], the CAB Thesaurus[11], or the USDA's National Agricultural Library NAL[12], the aggregated market data can be used for specific analyses for government or NGO purposes.

4.2 Linked Data Applications

We are currently building a client web application where we use the various Linked Data. This application will use the market data and exploit its links to GeoNames for displaying the market offerings on a map. The links to other dataset (IDS data, pluvial data, etc.) will also be exploited to provide the user with additional information about products or regions.

The application will allow the local NGO to perform various types of analysis based on market data that are useful on the basis of their educational programs. The application will aim to demonstrate the added value of the linked data approach through the re-use of and integration with existing market data from various sources with differing schemes and through the re-use of and integration with market data with publicly available knowledge from the web on agriculture and economics.

At the same time, we will continue to work on client applications for users in the developing regions themselves, focusing on voice-based access to the data. Within the VOICES project, (limited) TTS systems for smaller languages are being developed.

5 Discussion

We have presented the Linked Data version of the RadioMarché system, its data and the voice-based access. This system represents our first steps to brining Linked Data to producers and consumers in developing countries. We describe a demonstrator with locally produced Linked Data which provides rudimentary voice-based access, in addition to browser-based and Linked Data-application access.

Currently, the demonstrator is implemented on commercial-grade and University-provided web servers including the Voxeo Evolution platform, PURL servers and the VU University Amsterdam web server. The voice application is also only reachable through a Dutch local phone number or Skype access. To ensure sustainability of the Linked Data and the client applications, this infrastructure needs to be moved to the developing regions itself as much as possible. The Orange Emerginov platform [13] can provide the web server and voice browser technology needed for this infrastructure and include local Malian phone numbers. The Linked Data servers, voice-interfaces and client applications can be moved to this platform at testing or deployment time. A second option is entirely local. This version has the data and applications running

[10] http://aims.fao.org/website/AGROVOC-Thesaurus/sub
[11] http://www.cabi.org/cabthesaurus/
[12] http://agclass.nal.usda.gov/
[13] http://www.emerginov.org/

on a web-connected dedicated laptop that is be deployed locally. The voice channel is provided by a local voice browser and a GSM gateway (2N OfficeRoute) device connected to the laptop that allows phone calls to be handled by the system on the laptop.

As was discussed in Section 1.2, we aim to include the audio language resources to the Linked Data itself. We are currently gathering language snippets that act as audio labels for resources. These will be added to the data itself so that they can be interpreted by a voice browser directly.

Acknowledgements The authors would like to thank Mary Allen and Amadou Tangara of Sahel Eco. This research is partly funded by the European Union through the 7th Framework Programme (FP7) under grant agreement Num. 269954.

References

1. Guèret, C., Schlobach, S., de Boer, V., Bon, A., Akkermans, H.: "is data sharing the privilege of a few? bringing linked data to those without the web". Outrageous Ideas at International Semantic Web Conference (ISWC 2011). Jury award winning paper. 1st Place (2011)
2. de Boer, V., Leenheer, P.D., Bon, A., Gyan, N.B., van Aart, C., Guèret, C., Tuyp, W., Boyera, S., Allen, M., Akkermans, H.: Radiomarché: Distributed voice- andweb-interfaced market information systems under rural conditions. In: Accepted for publication in Proceedings of 24th International Conference on Advanced Information Systems Engineering, CAiSE'2012, Gdansk, Poland, 25 29 June 2012. (2012)
3. Domingue, J., Pedrinaci, C., Maleshkova, M., Norton, B., Krummenacher, R.: Fostering a relationship between linked data and the internet of services. In Domingue, J., Galis, A., Gavras, A., Zahariadis, T., Lambert, D., Cleary, F., Daras, P., Krco, S., Mller, H., Li, M.S., Schaffers, H., Lotz, V., Alvarez, F., Stiller, B., Karnouskos, S., Avessta, S., Nilsson, M., eds.: The Future Internet. Volume 6656 of Lecture Notes in Computer Science. Springer Berlin / Heidelberg (2011) 351–364
4. Agarwal, S.K., Jain, A., Kumar, A., Rajput, N.: The world wide telecom web browser. In: Proceedings of the First ACM Symposium on Computing for Development. ACM DEV '10, New York, NY, USA, ACM (2010) 4:1–4:9
5. ESOKO: Esoko. http://www.esoko.com/ (2011)
6. AppLab, G.: Google sms to serve needs of poor in uganda. http://blog.google.org/2009/06/google-sms-to-serve-needs-of-poor-in.html (2009)
7. Foundation, W.W.W.: Open government data. http://www.webfoundation.org/projects/ogd/ retrieved 14-03-2012 (2012)
8. 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)
9. Schreiber, G., Amin, A., van Assem, M., de Boer, V., Hardman, L., Hildebrand, M., Hollink, L., Huang, Z., van Kersen, J., de Niet, M., Omelayenko, B., van Ossenbruggen, J., Siebes, R., Taekema, J., Wielemaker, J., Wielinga, B.: Multimedian e-culture demonstrator. In: The Semantic Web - ISWC 2006, Athens, Georgia, volume 4273 of LNCS, pages 951-958, Winner Semantic Web Challenge 2006, Springer Verlag, November 2006 (2006)
10. W3C: Voice extensible markup language (voicexml) version 2.0. W3C Recommendation 16 March 2004 http://www.w3.org/TR/voicexml20/ (2004)