A Dialogue with Linked Data: Voice-based Access to Market Data in the Sahel

Victor de Boer a,∗ Nana Baah Gyan a Anna Bon b Wendelien Tuyp b Chris van Aart a Hans Akkermans a

a Department of Computer Science, VU University Amsterdam, The Netherlands
Email: {v.de.boer, n.b.gyan, c.j.vanaart, j.m..akkermans}@vu.nl
b Center for International Cooperation, VU University Amsterdam, The Netherlands
Email: {a.bon, w.a.m.tuyp}@vu.nl

Abstract. In this paper, we present our ongoing efforts to bring the benefits of 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 we have made RadioMarché data accessible as Linked Open Data and present a prototype demonstrator that provides 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.

Keywords: Voice Interface, Linked Data for Africa, Linked Market 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 cannot 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 access1, only 10% has access to the electricity network2, and only 26.2% is literate3. 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 investigations the implementation of Linked Data-backed solutions for the rural Sahel regions.

∗Corresponding Author


2http://www.developingrenewables.org/energyrecipes/reports/genericData/Africa/061129%20RECIPES%20country\%20info\%20Mali.pdf

3http://www.indexmundi.com/facts/indicators/SE.ATL.LITR.ZS Index Mundi 2011.
In Africa, mobile telephony has become the primary mode of telecommunication [2]. In 2006, an estimated 45 percent of Sub-Saharan African villages were covered by a mobile signal [3]. And in 2009, Africa showed the fastest rate of subscriber growth, introducing 96 million new mobile subscribers in a period of only twelve months [4]. The widespread availability of mobile phones and increasing level of coverage creates great opportunities for new services.

The contributions of this paper are the following.

- We describe of RadioMarché (RM), a MIS concept adapted for rural conditions in the African Sahel. Regarding the above mentioned challenges, RM is not dependent on Internet infrastructure, and has voice-based and sms-based interfaces. We describe the current status of deployment of this system in Mali, Africa.
- We have converted the market data gathered using RadioMarché and expose this data as Linked Open Data. We discuss the data, its model and opportunities for re-use. On a large scale, we deal with the issue of aggregation and management of distributed market data by adopting Linked Data approaches. We show how our design choices offers opportunities to link aggregated market information to datasets from other domains. The resulting “Web of Data” provides an open innovation platform to develop services with augmented reasoning capabilities for e.g. NGOs, governments, policy makers, traders and scientists.
- We propose a method of adding voice information to the linked data produced in this way through the use of pre-recorded audio files, accessible through the Web, thus allowing for re-use of this data in voice-based applications, relevant in the low-literate, low-infrastructure context.
- We finally describe a prototype voice application that uses this data. This demonstrator provides voice-based access to the RadioMarché linked data.

2. Related Work

Related work on voice technologies started in the 1930s in research on speech recognition. The first commercial deployments of voice-based services took place in the early 1970s. Major achievements on language recognition took place in the 1980s and 1990s, but this was mainly focused on English. While

Text-To-Speech and Speech Recognition are key in voice application development, the creation of the VoiceXML standard by the W3C Voice Browser group, in 1999, further facilitated the development of voice applications [5].

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 [6].

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 [7], 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 [8]. 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” [9]. 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. [10]. The SemanticXO is a system that connects rugged, low-power, low-cost robust small laptops for empowerment of poor communities in developing countries.

3. RadioMarché

Our efforts center around a Market Information System, RadioMarché [11], a web-based market information system developed within the VOICES project

4http://www.mvoices.eu
aimed at stimulating agricultural trade in the Sahel region. The RadioMarché system is a Market Information System (MIS) that supports local farmers of Non-Timber Forest Products (NTFPs) such as honey, tamarind and shea nuts. It was introduced under support of a local partner NGO, Sahel Eco, in the Tominian Area in Mali.

RadioMarché is used to distribute up-to-date market information via community radio in the area. A Sahel Eco staff member receives offerings from local farmer’s representatives in the form of an SMS text message, containing information about a product offer: quantity, quality, price, name of the seller, village, phone number, etc. The SMS information is entered manually into a web form. The information is then accessible for employees of local community radio stations through either the Web or through a first generation mobile-phone accessible voice interface.

For the voice interface, the market data is converted to audio files using a slot-and-filler Text to Speech (TTS) system[12]. This audio is produced in multiple languages, including “small” languages spoken in the region such as Bambara and Bomu. The generated audio is made accessible through an interactive voice response (IVR) system using the open standardized VoiceXML language. Radio station personnel can access the market information in their specific language using their phone and broadcast the audio to the public.

RadioMarché has been deployed from November 2011-November 2012 in the Tominian region in Mali. Feedback from this pilot phase was positive and showed that indeed the system was used to distribute market information in the region more effectively. In fact, the demand for honey offered became so large that the supply could not cope[13]. For this reason, the RadioMarché system is currently not operational until May 2013.

4. The Linked Market Data Demonstrator

In this section we describe a prototype demonstrator developed in parallel which exposes the market data gathered through the RadioMarché system using Linked Data approaches, so that new opportunities for product and service innovation in agriculture and other domains can be unleashed.

---

5http://www.saheleco.net
6http://www.w3.org/TR/voicexml20/
minian region to RDF triples using the XMLRDF tool\(^7\) of the ClioPatria semantic framework \([15]\). The XMLRDF tool allows us to convert XML documents to arbitrary RDF structures in an interactive way, using simple RDF graph rewriting rules. The conversion is run whenever the RadioMarché database is updated to ensure the database of the deployed version and the linked data store of our prototype are synchronized. The RDF triples are stored in a triple store at http://semanticweb.cs.vu.nl/radiomarche/.

We use PURLs for the resource URIs. The namespace chosen is http://purl.org/collections/w4ra/radiomarche/. An HTTP request to these PURL URIs is redirected to the ClioPatria server. 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 or the raw RDF triples describing the resource depending on the accept header in the HTTP request. For example rm:village_Samoukuy/ shows all information about the Samoukuy village. A SPARQL endpoint is also provided at http://semanticweb.cs.vu.nl/radiomarche/sparql/\(^8\).

As of March 2013, 90 market offerings are in the triple store. These market offerings have been done by 17 different farmers, living in 16 different villages spread across 6 regional “zones”. The market offerings contain the quality, quantity and type of the product the price and contact information. Furthermore, offerings are grouped in so-called “communiqués”, designed to be sent to local radio’s for broadcast. In total, the market data consists of 1,952 triples.

We have manually added a number of mapping links:

- Zones and villages are mapped to GeoNames\(^9\) concepts. Not all villages identified in the GeoNames dataset, speaking to the remote nature of the data.

- The product types were mapped to the Agrovoc thesaurus\(^10\) which contains in up to 21 languages covering topics related to food, nutrition, agriculture, fisheries, forestry, environment and other related domains.

- Some zones and villages as well as products are mapped to DBpedia\(^11\) resources.

To show the added value of linking the locally produced market data to these external sources on the Web of Data, we have developed prototype mashup application that uses these links to provide support to local NGOs for monitoring the local markets and help them sharing this information. This application could be used by the NGOs to plot local product offerings on an interactive map (using GeoNames) and presents external product information (from Agrovoc) to the user[16].

4.3. Linked Voice Data

The RDF model allows for a single resource, identified by a URI (eg. rm:shea_nuts) to have multiple language labels (eg. “Shea Nuts”@en and “Amande de Karité”@fr). For widely spoken world languages, such as English, French or Chinese, good TTS systems are available, that could be used by off-the-shelf voice-based systems to pronounce a product name in that language. However, for smaller languages, these TTS systems do not exist. In our case, the Bambara and Bomu languages of the RadioMarché stakeholders lack such systems. In this case, augmenting the data with voice labels increases the opportunities for re-using this market data for voice-application. Since this concerns limited amount of data within a closed domain, it is feasible to record voice labels for the data items.

We propose voice labels, that are complementary to rdfs:label constructs and link web-accessible wave files, identified through URIs to RDF resources. For this purpose, we introduce the speakle:voicelabel properties. The current namespace for the speakle schema is http://purl.org/collections/w4ra/speakle. To allow for multi-lingual voice labels, we furthermore define language-specific voice label sub-properties (speakle:voicelabel_en, speakle:voicelabel_fr etc.). Rather than have a literal as range, the voice label property points to a web-accessible wave file. Voice applications can fetch these audio files when needed and insert them into the voice user interface as needed. The speakle schema is currently very limited but can be extended to allow for example for audio labels of different speakers. In Figure 4.3, we show how in the case of our market data audio and literal labels can be mixed, for different languages.

\(^{7}\)http://semanticweb.cs.vu.nl/xmlrdf/
\(^{8}\)An interactive query environment is provided at http://semanticweb.cs.vu.nl/radiomarche/flint/
\(^{9}\)http://www.geonames.org/
\(^{10}\)http://www.fao.org/agrovoc/
\(^{11}\)http://en.dbpedia.org/
As a proof of concept, for our Linked Market Data demonstrator, we recorded a number of audio labels in Dutch and English, and added them to the Linked Data using the voice label triples.

4.4. Voice-based access to Linked Data

As presented in the previous sections, the Linked Market Data can be browsed through the web as well as be re-used in mashup applications. 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. As a proof-of-concept, we have implemented a voice-based application to the linked market data as described in the previous section. The voice application is built using VoiceXML, the industry standard for developing voice applications. The application allows limited browsing of the latest market offerings. We have implemented a partial bi-lingual interface, in English and Dutch, using the voice labels in the voice interface.

The prototype voice application is running on the Voxeo Evolution platform. This 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 +99000936 9996162208 and the local (Dutch) phone number +31208080855. When any of these numbers is called, the voice application accesses a VoiceXML document. This document defines the dialogue structure for the application. In the current demonstrator, the caller is presented with two options, to browse the data by product or region, or to listen to the latest offerings. The caller presses the code on his or her keypad (Dual Tone Multi-Frequency or DTMF). The voice application interprets the choice and forwards the caller to a new voice menu. 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 passed as a HTTP GET variable.

Based on the choice, a in the PHP document a SPARQL query is constructed. This SPARQL query is then passed to the Linked Market Data server, which returns the appropriate results in the correct language.

For a product query, all (recent) offerings about that product are returned, including the links to the wave files of the quantity, unit, contact information etc. For each result the PHP code inserts the URLs of the voice resources into the VoiceXML code at the appropriate spot. In this way, a sequence of audio files is played to the caller, providing him or her with the requested information. Figure 3 shows a snippet of the VoiceXML code.

![Fig. 3. Example VoiceXML code snippet, generated by the PHP code, using the results from a SPARQL query. For brevity, the URLs of the wave files are abbreviated (...). Note that the introduction “The following...”, “offers” and “for” are to be interpreted by the (in this case English) TTS. These utterances are considered part of the application and it would be the responsibility of the application developer rather than the (linked) data provider to provide audio versions for them.](image)

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 user interfaces- 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.

5. Discussion and Current Work

The efforts above represent first efforts into exposing locally produced data in developing regions as

---

12 Languages spoken by authors

13 http://evolution.voxeo.com
linked data. We have described a small prototype voice application that provides access to this data. Currently, to allow for easy testing, this access is in English and Dutch only. However, for the RadioMarché system, a great number of audio recordings -used in the slot-and-filler TTS- are available that correspond to the data items in the Linked Market Data. We can use these to add Bambara and Bomu voice labels to the linked data. We are currently planning on realizing this full integration.

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\(^{14}\) 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 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 OfficeR-out) device connected to the laptop that allows phone calls to be handled by the system on the laptop.

Finally, the market data is not the only data relevant to the development context that we want to expose as linked data. In a number of related projects, we are producing linked data sets that 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 meet-

\(^{14}\)http://www.emerginov.org/
ing 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.

**Pluvial Data.** We have developed a crowdsourcing platform to transform photocopied data about rainfall in the Bankas area in Mali to Linked Open Data[17]. 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 and IATI Data.** The Institute for Development Studies recently published an API exposing more than 30,000 publications about development research[15]. We have developed a wrapper around the IDS API to expose its content as high quality Linked Data, enriching it with connections to other Linked Data datasets[16]. These include both general datasets such as DBPedia or GeoNames as well as datasets with information from developing countries that are currently being realized. The same wrapper also exposes data from the International Aid Transparency Initiative (IATI)[17], which lists administrative information about development projects. Our aim is to link this more general development-related data to the results of those development efforts in the field, thereby increasing the value at both levels.

By integrating the Linked Market data with these and other datasets, we can create a Web of Data that is relevant to as well as accessible for people living in developing areas, bridging the digital divide.

Acknowledgements

The authors would like to thank Mary Allen and Amadou Tangara of Sahel Eco, Stephane Boyera, Max Froumentin and Aman Grewal of the Web Foundation and the farmers and radio employees in Mali for their collaboration. This research is partly funded by the European Union through the 7th Framework Programme (FP7) under grant agreement Num. 269954.

References


