Semantic Web 0 (0) 1
IOS Press

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

Editor(s): Christophe Guéret, Data Archiving and Networked Services (DANS), KNAW; Stephane Boyera, SBC4D; Mike Powell, IKM Emergent; Martin Murillo, Data Connectivity Initiative – IEEE
Solicited review(s): Michalis Vafopoulos, National Technical University of Athens, Greece; Martin Murillo, Data Connectivity Initiative – IEEE; Ivana Marenzi, L3S Research Center, Hannover, Germany

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.tuyp}@vu.nl

Abstract. The Linked Data movement has facilitated efficient data sharing in many domains. However, people in rural developing areas are mostly left out. Lack of relevant content and suitable interfaces prohibit potential users in rural communities to produce and consume Linked Data. In this paper, we present a case study exposing locally produced market data as Linked Data, which shows that Linked Data can be meaningful in a rural, development context. We present a way of enriching the market data with voice labels, allowing for the development of applications that (re-)use the data in voice-based applications. Finally, we present a prototype demonstrator that provides access to this linked market data through a voice interface, accessible to first generation mobile phones.

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 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 [1]. 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]. Furthermore, there is a need for knowledge sharing systems to deal with specific use cases and content relevant to people in the target communities. Locally relevant content can be an incentive for people to access and use the knowledge sharing systems [1].

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 [2], while the Web of Documents has been around for 20 years, as engineers

*Corresponding Author

[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

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 unbridgeable gap, we should consider the underprivileged majority and these barriers as we design Linked Data architecture, describe use cases and provide access to that Linked Data.

1.1. Mobile technologies

One way of overcoming the infrastructural and literacy barriers is to turn to simple mobile phones, GSM architecture and voice technologies to build knowledge sharing services. In Africa, mobile telephony has become the primary mode of telecommunication [3]. In 2006, an estimated 45 percent of Sub-Saharan African villages were covered by a mobile signal[4]. And in 2009, Africa showed the fastest rate of subscriber growth, introducing 96 million new mobile subscribers in a period of only twelve months [5]. The widespread availability of mobile phones and increasing level of coverage creates great opportunities for new services.

1.2. Linked Data for knowledge sharing in rural areas in developing countries

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 [6]. 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. In Section 4.2, we present an example of a prototype application reusing linked market data.

An additional advantage is that Linked Data is well-suited for localization efforts. Linked Data is very much suited for dealing with multilinguality 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 agnostic", which suits our purpose of multilingual and voice-based access well. In Section 4.3, we elaborate on our efforts to attach 'voice labels' to the data.

The combination of Linked Data, mobile services and voice technology allows us to integrate and share locally relevant information, while at the same time allowing local stakeholders to access the data using available hardware and voice interfaces. In this paper, we describe a prototype solution based on Linked Data for a specific use case in the rural Sahel region that allows voice-based access to the data.

1.3. Contributions

The contributions of this paper are the following.

- We describe a knowledge sharing use case in Mali, Africa and a Market Information System (MIS) adapted to the local context, RadioMarché (RM). 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..
- The market data gathered using RadioMarché have been converted to and exposed as Linked Open Data. In Section 4.1, we discuss the data, its model and opportunities for re-use. We show how our design choices offers opportunities to link aggregated market information to datasets from other domains. The resulting data provides an open innovation platform to develop services with augmented reasoning capabilities for local entrepeneurs, 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 also describe a partial implementation of this voice data for the RadioMarché Linked Data case.
- We have developed a prototype voice application that uses this data. This proof-of-concept demonstrator provides voice-based access to the RadioMarché linked data. This demonstrator is described in Section 4.4.

2. Related Work

We can find many efforts towards the usage of mobile technologies in an ICT for Development (ICT4D)

context. In 2013 the International Telecommunication Union (ITU) organized a forum discussing how mobile and broadband connectivity can be a tool in global economic development [4]. There are many international research initiatives and practitioners turning to mobile services for information sharing in rural developing areas. Examples include SPIDER [5], IDRC [6] or the Nethope network of ICT4D practitioners [7]. Success stories of such mobile services include studies in fishing villages in India [7], grain markets in Niger [8] or crop markets in Uganda [9].

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 [10].

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 [11].

Several automated Market Information Systems (MIS) have been developed and built to support farmers and agricultural trade in developing countries. One of the well-known systems is ESOKO [8], 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 [12]. 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" [13]. This effort focuses on opening government data in developing countries such as Ghana. Our data is initially designed to be produced and consumed 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. [14]. 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 Market Information System

Our efforts center around, RadioMarché [15,16], a web-based information system developed within the VOICES project[9] aimed at stimulating agricultural trade in the Sahel region. The RadioMarché system is a 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[10], 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 [17]. 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[11] language. Radio station personnel can access the market information in their specific lan-

[4] http://www.itu.int/en/wtpf-13/
[5] https://spidercenter.org/
[6] http://www.idrc.ca/
[7] http://nethope.org/
[8] http://www.esoko.com/

[9] http://www.mvoices.eu
[10] http://www.saheleco.net
[11] http://www.w3.org/TR/voicexml20/

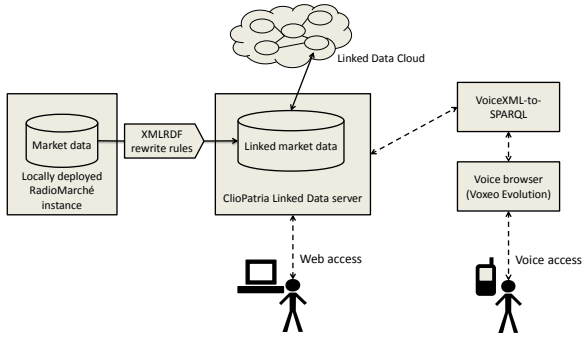


Fig. 1. Schematic representation of the linked market data prototype demonstrator.

guage 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 [16]. For this reason, the RadioMarché system is at the time of writing not collecting and disseminating new offerings. However, Sahel Eco is currently looking to restart the service, embedding the service in other efforts that stimulate the value-chain. One key lesson here is that even a successful ICT deployment is not sustainable if further up or down the value-chain the requirements created by the new situation can not be met.

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.

4.1. The linked market data

The high-level setup of the Linked Data prototype demonstrator and how it relates to RadioMarché is shown in Figure 1. To obtain the Linked Data, we translate a copy of the up-to-date market information from the RadioMarché prototype deployed in the Tominian region to RDF triples using the XMLRDF tool[12] of the ClioPatria semantic framework [18]. 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[13].

We utilize PURLs for the resource URIs with an appropriate namespace chosen[14]. 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[15].

As of March 2013, 90 market offerings are in the triple store. These market offerings are made 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 as well as 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[16] 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[17] which contains in up to 21 languages covering topics related to food, nutrition, agriculture, fisheries, forestry, environment and other related domains.

[12] `http://semanticweb.cs.vu.nl/xmlrdf/`

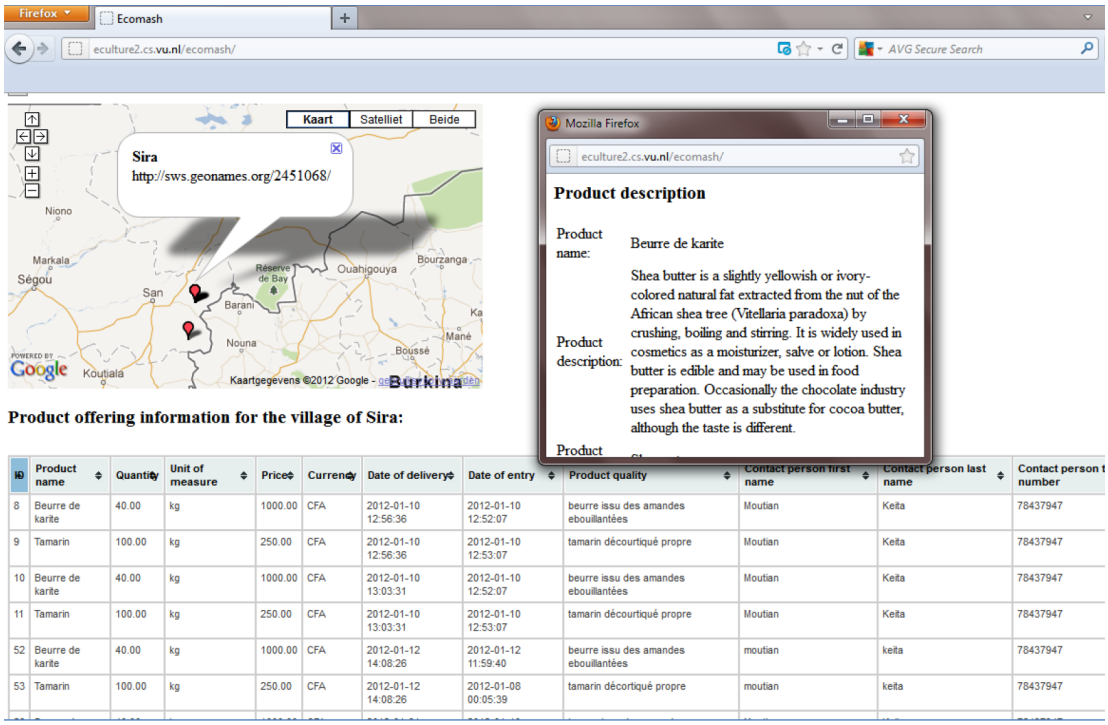
[13] The triple store is available at `http://semanticweb.cs.vu.nl/radiomarche/`

[14] `http://purl.org/collections/w4ra/radiomarche/`

[15] `http://semanticweb.cs.vu.nl/radiomarche/sparql/`, an interactive SPARQL query environment is provided at `http://semanticweb.cs.vu.nl/radiomarche/flint/`

[16] `http://www.geonames.org/`

[17] `http://www.fao.org/agrovoc/`



ID	Product name	Quantity	Unit of measure	Price	Currency	Date of delivery	Date of entry	Product quality	Contact person first name	Contact person last name	Contact person t number
8	Beurre de karite	40.00	kg	1000.00	CFA	2012-01-10 12:56:36	2012-01-10 12:52:07	beurre issu des amandes ebouillantées	Moutian	Keita	78437947
9	Tamarin	100.00	kg	250.00	CFA	2012-01-10 12:56:36	2012-01-10 12:53:07	tamarin décourtiqué propre	Moutian	Keita	78437947
10	Beurre de karite	40.00	kg	1000.00	CFA	2012-01-10 13:03:31	2012-01-10 12:52:07	beurre issu des amandes ebouillantées	Moutian	Keita	78437947
11	Tamarin	100.00	kg	250.00	CFA	2012-01-10 13:03:31	2012-01-10 12:53:07	tamarin décourtiqué propre	Moutian	Keita	78437947
52	Beurre de karite	40.00	kg	1000.00	CFA	2012-01-12 14:08:26	2012-01-12 11:59:40	beurre issu des amandes ebouillantées	moutian	keita	78437947
53	Tamarin	100.00	kg	250.00	CFA	2012-01-12 14:08:26	2012-01-08 00:05:39	tamarin décortiqué propre	moutian	keita	78437947

Fig. 2. Screenshot of the "Ecomash" prototype application. This application uses linked external information from GeoNames, DBPedia and Agrovoc to display data on a map and provide background information [19].

- Some zones and villages as well as products are mapped to DBpedia[18] resources.
- To facilitate reuse of this e-business data, we have therefore established a partial mapping from our schema to the GoodRelations vocabulary [20]. In total 7 classes and 9 properties are mapped to GoodRelations classes and properties.
- Similarly, we have added 5 mappings from RM classes and properties to the FOAF vocabulary[19].

4.2. Ecomash

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, Ecomash. This application was based on requirements put forward by our local NGO partner. Figure 2 shows a screenshot of this mashup application. This application is designed to be used by this and other NGOs to monitor local markets. Local product offerings can be filtered in an interactive table or browsed via an interactive map (using the GeoNames links) and external product information (from the DBPedia and Agrovoc links) are presented to the user[19].

4.3. Linked voice data

RDF 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 off-the-shelf TTS systems are available, which can be used by voice-based systems to utter 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-applications. Since this concerns a 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`

[18] `http://en.dbpedia.org/`
[19] `http://xmlns.com/foaf/spec/`

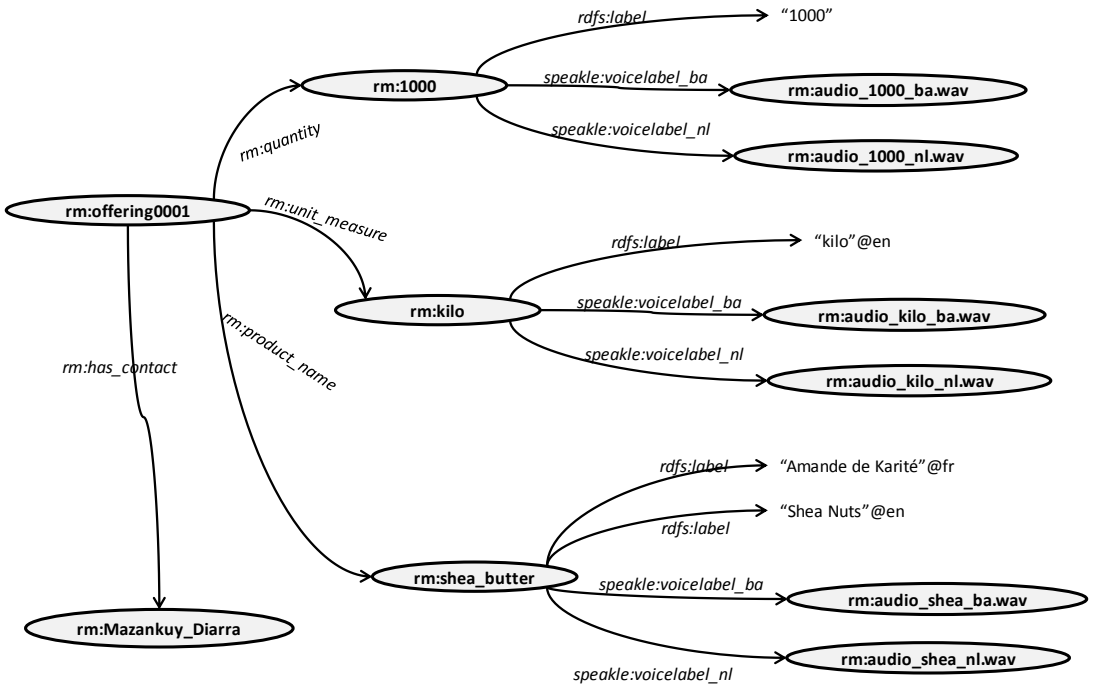


Fig. 3. Example RDF graph snippet of linked market data in the triple store. Resources have URIs and are represented using ellipses, typed relations between resources and to literal labels are represented using arrows. The namespace `http://purl.org/collections/w4ra/radiomarche/` is abbreviated to `rm:`. The figure shows how web-accessible multilingual audio files are related market data resources. Note that only a small part of the market data related to offerings and product types is shown.

properties[20]. To allow for multilingual voice labels, we furthermore define language-specific voice label subproperties (`speakle:voicelabel_en`, `speakle:voicelabel_fr` etc.). Rather than have a literal as range, the voice label property points to an audio web resource such as a 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 3, we show how in the case of our linked market data audio and literal labels can be mixed, for different languages. As a proof of concept, for our Linked Market Data demonstrator, we recorded a number of audio labels in Dutch and English[21] 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 RadioMarché linked market data can be browsed through a web interface or can be re-used in mashup applications. However, as stated, our goal is to provide a voice-based interface that allows market information access for users having a first-generation mobile phone. As a proof-of-concept, we have implemented a voice-based interface to the linked market data. The voice application is built using VoiceXML [10], the industry standard for developing voice applications. The application allows limited browsing of the latest market offerings. We have implemented a partial bilingual interface, in English and Dutch, using the voice labels in the voice interface.

The prototype voice application is running on the Voxeo Evolution platform[22]. This platform includes a voice browser, which is able to interpret VoiceXML documents, includes (English) TTS and provides a

[20]The current namespace for the speakle schema is `http://purl.org/collections/w4ra/speakle`

[21]Languages spoken by the authors

[22]`http://evolution.voxeo.com`

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 offering. 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. The application then accesses a PHP document on the remote server, the choice is passed as a HTTP GET variable.

In the PHP document a SPARQL query corresponding to the user choice 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 resource 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 4 shows a fragment of the VoiceXML code.

Of course, the current method of accessing the data is only one of many possible actions. In a different voice application a caller could 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.

```
<vxml version = "2.1" >
 <property name="inputmodes" value="dtmf" />
 <form id="result">
  <block>
   <prompt>
    These are the top three current
    offerings for Tamarin
    <break time="0.5s"/>
    <audio src="http://.../AndiaKalakodio_en.wav"/>
    offers
    <audio src="http://.../450_00_en.wav"/>
    <audio src="http://.../Kg_en.wav"/> for
    <audio src="http://.../250_00_en.wav"/>
    <audio src="http://.../CFA_en.wav"/>
    <break time="0.5s"/>
    ...
    <break time="0.5s"/>
   </prompt>
   <goto next="mytest.xml"/>
  </block>
 </form>
</vxml>
```

Fig. 4. Example VoiceXML code snippet, generated by the PHP code, using the results from a SPARQL query. For brevity, the URLs of the audio 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.

5. Discussion and work in progress

The previous sections describe first efforts into exposing locally produced data in rural areas in developing countries as 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. Any missing items, such as the application-specific utterances would have to be recorded locally.

5.1. Sustainability

As stated in Section 3, the RadioMarché service is currently online but not accepting or distributing new offerings due to issues in different parts of the value chain. Sahel Eco is currently looking at restarting the service, which would result in a steady flow of new market information into the database and therefore in the linked data store.

To ensure sustainability, we are currently transferring the hardware and software that runs the service to a local business partner[23]. Local ownership and control is one of the key issues for the sustainability of rural ICT projects [21]. Within the VOICES project,

[23] http://yeleman.com/

multiple business models have been developed to analyze the financial sustainability of the RadioMarché project [22].

One issue concerns the costs of phone calls for accessing the voice interface. Interviews with stakeholders confirm a willingness to pay for phone calls to a system, given that some economic benefit is gained from the call. However, we are looking at ways of limiting these costs. One opportunity is allowing for access to data via (cheaper) SMS text messages. Simple text queries can be processed by the MIS and replies can be presented in SMS text messages. An added benefit is that the answer is stored on the user's device. However, in our specific case study, most users have a low literacy level and are not comfortable using text messages. The arrival of higher-generation mobile phones (i.e. feature phones) allows the development of text-to-voice applications that can generate voice messages out of the text. This would overcome the literacy barrier, while still cutting costs. However, currently such phones are not available in our case study context.

Another way of reducing costs for end users is having other stakeholders reimburse costs made for calling. We are currently investigating various value propositions which involve different local stakeholders including local entrepreneurs, NGOs, farmers, potential buyers and community radio stations broadcasting the market data.

Currently, the linked market data and the voice interface demonstrator are still maintained using 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. Sustainability of the linked data and the client application is currently ensured through the stability of the Linked Data platform at the University and the ability to convert new data as it is gathered in the RadioMarché system. Even beyond the VOICES project, this service will be maintained at the university. However, to ensure sustainability in the long run, this infrastructure would also have to be moved to the rural regions itself as much as possible. The Orange Emerginov platform[24] 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 OfficeRoute) device connected to the laptop that allows phone calls to be handled by the system on the laptop.

Before deploying a next version of the system, user evaluation of the voice interface and the system as a whole is necessary.

5.2. Related linked data sets under production

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 we have developed a voice system for a second use case: a meeting scheduling system [16]. This system "Tabale" provides 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 [23]. 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 correlations 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[25]. 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[26]. These include both general datasets such as DBPedia or GeoN-

[24] http://www.emerginov.org/

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

[26] http://api2lod.appspot.com/

ames as well as datasets with information from developing countries that are currently being realized.

IATI data We have also produced a linked data from the International Aid Transparency Initiative (IATI)[27], which lists transparency data about development projects. The data is linked to external resources to reuse and visualize this data. Our aim is to link this more general development-related data to the results of those development efforts in the field, through for example funding organizations. By linking the data that describes project results to the IATI data, NGO's can increase their own transparency and accountability. At the same time, we are investigating voice-based access to this transparency linked data using the voice labels as described here, which would allow people in rural areas to query the aid data that concerns them.

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.

6. Conclusions and further work

In this paper, we have described a Linked Data set derived from a Market Information System currently deployed in Mali, Africa. This Linked Data set is up-to-date and available for re-use. This dataset represents data that is relevant to the different stakeholders in the specific region and therefore shows that Linked Data could be valuable within that region. We have established links to related datasets and vocabularies allowing for new applications that use the integrated data. We show one such mashup application, the Ecomash prototype. This case study shows that indeed meaningful Linked Data can be produced and used in a rural, development context. We show that this data can be enriched by establishing links to external sources. We also describe how this data can be reused in new and useful ways, including through a prototype mashup application.

We also show how the linked data paradigm allows the market data to be easily expanded with audio resources, allowing for development of applications that (re-)use the linked data in voice-based applications. We have also developed one such voice-based application allowing access to the linked market data.

In terms of this specific case, future work includes the further development and user testing of the application. For this, further localization is needed and requires adding French, Bambara and Bomu audio labels. Using this, we can recreate the current RadioMarché voice user interface using the (Voice) Linked Data. The choice for English and Dutch makes that the proof-of-concept is not directly deployable in rural areas in Mali as is. However, since the RadioMarché system's audio fragments are nearly identical, we actually have gathered most of the French, Bambara and Bomu audio fragments over the course of the project. We are currently converting some of these audio resources and linking them to the data as voice labels in the manner described in 4.3. As far as gathering new voice fragments is concerned, for specific domains scripts have to be constructed that cover all of the data items. As an illustration of the effort needed, we have made an example of such a script used in the RadioMarché MIS available for download[28]. Furthermore, future work to show the added benefit of the Linked Data approach is to also develop different voice applications that reuse these audio linked market data voice labels.

We are currently also looking at adding voice labels to different linked data, for example cultural heritage information in the Netherlands. This could be used in a voice-based application aimed at -for example- users with accessibility issues.

We are also investigating methods of gathering voice fragments using crowdsourcing initiatives [24]. Crowdsourcing can be a good way of producing large amounts of voice fragments for relatively small cost. This would allow for larger-scale projects and make further progress towards a voice accessible Web of Data.

Acknowledgements

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

[27] http://iatistandard.org/

[28] http://www.few.vu.nl/~vbr240/voices/recording_script.pdf

References

[1] Stephane Boyera, W3C. Mobile mobile web for social development roadmap. W3C Interest Group Note 17 November 2009 http://www.w3.org/TR/2009/NOTE-mw4d-roadmap-20091208/, 2009.

[2] Christophe Guéret, Stefan Schlobach, Victor de Boer, Anna Bon, and Hans Akkermans. Is data sharing the privilege of a few? Bringing linked data to those without the web. In *Proceedings of the International Semantic Web Conference (ISWC 2011) "Outrageous ideas" track, Best paper award*, pages 1–4, 2011.

[3] UNCTAD. Science and technology for development: the new paradigm of ICT. Information Economy Report 2007–2008, United Nations Conference on Trade and Development – http://unctad.org/en/Docs/sdteecb20071_en.pdf, 2007.

[4] ITU. Telecommunications / ICT markets and trends in Africa. International Telecommunications Union report – http://www.itu.int/ITU-D/ict/statistics/material/af_report07.pdf, 2007.

[5] The Economist. Mobile marvels. Special report on telecoms in emerging markets. The Economist, September 2009 – http://www.economist.com/node/14483896, 2009.

[6] John Domingue, Carlos Pedrinaci, Maria Maleshkova, Barry Norton, and Reto Krummenacher. Fostering a relationship between linked data and the internet of services. In John Domingue, Alex Galis, Anastasius Gavras, Theodore Zahariadis, Dave Lambert, Frances Cleary, Petros Daras, Srdjan Krco, Henning Müller, Man-Sze Li, Hans Schaffers, Volkmar Lotz, Federico Alvarez, Burkhard Stiller, Stamatis Karnouskos, Susanna Avessta, and Michael Nilsson, editors, *The Future Internet*, volume 6656 of *Lecture Notes in Computer Science*, pages 351–364. Springer Berlin / Heidelberg, 2011.

[7] Robert Jensen. The digital provide: Information (technology), market, performance, and welfare in the south Indian fisheries sector. *The Quarterly Journal of Economics, Issue 3*, CXXII(879–924):879–924, August 2007.

[8] Jenny C. Aker. Does digital divide or provide? The impact of cell phones on grain markets in Niger. Working Paper 154. New York: Center for Global Development, 2008.

[9] Megumi Muto and Takashi Yamano. The impact of mobile phone coverage expansion on market participation: Panel data evidence from Uganda. *World Development*, 37(12):1887–1896, 2009.

[10] W3C. Voice extensible markup language (VoiceXML) version 2.0. W3C Recommendation 16 March 2004 http://www.w3.org/TR/voicexml20/, 2004.

[11] Sheetal Agarwal, Arun Kumar, Amit Anil Nanavati, and Nitendra Rajput. The world wide telecom web browser. In *Proceedings of the First ACM Symposium on Computing for Development*, ACM DEV '10, pages 4:1–4:9, New York, NY, USA, 2010. ACM.

[12] Rachel Payne. Google SMS to serve needs of poor in uganda. Google Applab blog post – http://blog.google.org/2009/06/google-sms-to-serve-needs-of-poor-in.html, 2009.

[13] World Wide Web Foundation. Open government data. http://www.webfoundation.org/projects/ogd/ retrieved 14-03-2012, 2012.

[14] Christophe Guéret and Stefan Schlobach. SemanticXO : connecting the XO with the world's largest information network. In Jim James Yonazi, Eliamani Sedoyeka, Ezendu Ariwa, and Eyas El-Qawasmeh, editors, *Proceedings of the First International Conference on eTechnologies and Networks for Development, ICeND2011*, volume 171 of *Communications in Computer and Information Science, Springer LNCS*, pages 261–275. Springer Berlin Heidelberg, 2011.

[15] Victor de Boer, Pieter De Leenheer, Anna Bon, Nana Baah Gyan, Chris van Aart, Christophe Guéret, Wendelien Tuyp, Stephane Boyera, Mary Allen, and Hans Akkermans. RadioMarché: Distributed voice- and web-interfaced market information systems under rural conditions. In Jolita Ralyté, Xavier Franch, Sjaak Brinkkemper, and Stanislaw Wrycza, editors, *Advanced Information Systems Engineering*, volume 7328 of *Lecture Notes in Computer Science*, pages 518–532. Springer Berlin Heidelberg, 2012.

[16] Nana Baah Gyan, Victor de Boer, Anna Bon, Chris van Aart, Hans Akkermans, Stephane Boyera, Max Froumentin, Aman Grewal, and Mary Allen. Voice-based web access in rural Africa, 2013.

[17] Etienne Barnard, Marelie H. Davel, and Gerhard B. van Huyssteen. Speech technology for information access: a South African case study. In *Proceedings of the AAAI Spring Symposium on Artificial Intelligence for Development (AI-D), Palo Alto, California, March 2010*, pages 8–13, 2010.

[18] Guus Schreiber, Alia Amin, Lora Aroyo, Mark van Assem, Victor de Boer, Lynda Hardman, Michiel Hildebrand, Borys Omelayenko, Jacco van Osenbruggen, Anna Tordai, Jan Wielemaker, and Bob Wielinga. Multimedian e-culture demonstrator. In *The Semantic Web – ISWC 2006, Athens, Georgia, Winner Semantic Web Challenge 2006*, volume 4273 of *Lecture Notes in Computer Science*, pages 951–958. Springer Verlag, November 2006, 2006.

[19] Henk Kroon. Ecomash: Linked market data for Sahel Eco. M.Sc. Thesis VU University Amsterdam http://www.few.vu.nl/~vbr240/teaching/Henk Kroon – Master thesis.pdf, 2012.

[20] Martin Hepp. Goodrelations: An ontology for describing products and services offers on the web. In Aldo Gangemi and Jérôme Euzenat, editors, *Knowledge Engineering: Practice and Patterns*, volume 5268 of *Lecture Notes in Computer Science*, pages 329–346. Springer Berlin Heidelberg, 2008.

[21] Caroline Pade, Brenda Mallinson, and David Sewry. An exploration of the critical success factors for the sustainability of rural ICT projects – the Dwesa case study. In *Information Systems Development*, pages 339–352. Springer, 2009.

[22] Hans Akkermans, Chris van Aart, Victor de Boer, Nana Baah Gyan, Anna Bon, Wendelien Tuyp, and Amadou Tangara. m-Agro knowledge sharing field pilot. Final evaluation. VOICES project deliverable document D5.4 (unpublished) http://mvoices.eu/sites/default/files/VOICES-D5.1-v1.1-VUA_WF_SE-16Dec2011-small.pdf, 2013.

[23] Victor de Boer, Michiel Hildebrand, Lora Aroyo, Pieter De Leenheer, Chris Dijkshoorn, Binyam Tesfa, and Guus Schreiber. Nichesourcing: Harnessing the power of crowds of experts. In Annette ten Teije, Johanna Volker, Siegfried Handschuh, Heiner Stuckenschmidt, Mathieu d'Acquin, Andriy Nikolov, Nathalie Aussenac-Gilles, and Nathalie Hernandez, editors, *Knowledge Engineering and Knowledge Management*, volume 7603 of *Lecture Notes in Computer Science*,

pages 16–20. Springer Berlin Heidelberg, 2012.

[24] Rokhsareh Nakhaie. Designing the open voice platform: A multi-lingual, crowd-sourced, and open platform to enable the creation of voice based applications for use in internet-deprived environments. M.Sc. Thesis VU University Amsterdam http://www.few.vu.nl/∼vbr240/teaching/Rokhsareh_Nakhaie_Master_thesis.pdf, 2012.