



Let's “Downscale” Linked Data

Christophe Guéret • *Data Archiving and Networked Services*

Victor de Boer and Stefan Schlobach • *Vrije Universiteit Amsterdam*

Open data policies and linked data publication are powerful tools for increasing transparency, participatory governance, and accountability. A closer look at linked data technologies, however, proves that their design and deployment exclude the majority of the world's population. It will take small but fundamental changes to bridge this gap.

The Internet has revolutionized the world and, with it, the way many of us do business, collect information, and communicate. Linked data is expected to advance information exchange to a new level by connecting machine-readable information to facilitate data interoperability and the generic reuse of knowledge. Many consider the impact of the Internet, along with the World Wide Web and the Web of Data, to be global, but nothing could be more untrue.

Nigel Shadbolt and Kieron O'Hara recently discussed the importance of governmental open data portals and how linked data can improve their usefulness and usability.¹ Contrary to what the open data community commonly believes, however, these Web-based platforms' societal impact is limited. Worldwide, an estimated 2 billion individuals are believed to be “online” and can thus access Web-hosted data platforms. Looked at in the context of a world population of 7 billion, heading toward 10, this proportion is rather low. Across and often within countries, a digital divide separates those who have access to information technologies from those who do not. To address this bottom of the pyramid and bridge the divide, we must contextualize ICT solutions, and preserve local culture and natural diversity.² Linked data technologies' flexibility and open nature make them a perfect fit for sharing contextualized and multilingual knowledge using open standards. Tools that leverage linked data could significantly contribute to improving data flows in less privileged parts of the world. Unfortunately, current

technologies can reach only a tiny fraction of the global population owing to the infrastructural reality in substantial areas of the world.

The good news is that, from a research viewpoint, we need to put only a few items on the linked data community's agenda to ensure that having access to these technologies becomes the norm, not a privilege.

The Limits to Data Sharing

In sharp contrast to the formal economy that drives developed countries, developing countries are often characterized by an *informal economy*, which has the following features: low requirements in terms of investment, small scale of operations, hands-on skills, and labor-intensive methods. Such an economy is often exemplified by sole street traders, artisans, subsistence farmers, and micro enterprises.³ To support such people's data needs, linked data must sustain direct information flow among street traders, artisans, and farmers more than top-down communication via governmental data portals.

One example is the wealth of knowledge and data used in farmer-managed natural regeneration (FMNR) techniques⁴ to fight desertification, and re-green soils. For most people in the world, radio and mobile phones are currently the leading ICT channels. When connectivity is missing for computer-enhanced communication, physical transportation (trucks and motorcycles) takes the Internet's role by asynchronously moving data from shop to shop.⁵ Rather than being replaced,

such existing communication channels must be enhanced and scaled up to share more information more efficiently. But several factors limit the ambition for opening and sharing in underprivileged parts of the world:⁶

- **Technology.** The lack of connectivity and electricity, as well as device and connection costs, are limiting the adoption and usage of new technologies.
- **Capacity.** A lack of time and resources limits participation in data sharing processes. Low education levels, low capacity to interpret data, and illiteracy are also problematic.
- **Relevance.** Power balance, culture, apathy, lack of incentives, lack of interest, and disempowerment are all threats to having citizens engage in data sharing.

To have real global impact, linked data must move beyond governmental data portals and large-scale systems to be usable on small, affordable hardware deployed in various connectivity contexts; accessible to individuals with varied cultural backgrounds or literacy levels; and relevant and directly useful to the target public they aim to empower.

Acting on Research Agendas

Linked data advocates can no longer ignore the majority of the world's population.⁷ When designing a new linked-data-based solution, we must ask ourselves several questions: Will it run on small hardware and with an unstable Internet connection? Can illiterate individuals use it? Who will be contributing data and system components? A critical look at linked data is necessary to adapt it and make it globally usable right now. We coined the term "downscaling" to refer to fitting linked data into weaker infrastructure, providing multimodal interfaces, and improving contextualized data sharing.

Limited Infrastructures

Web 2.0 technologies and cloud computing let us address data-sharing problems in a unique way: a server operates on the cloud, and the user community remotely queries it thanks to fast, stable, and reliable Internet connections. This interaction happens through interfaces and APIs designed to fit the target population's global needs.

Most linked data systems make no exception to this. The content of the Linked Open Data cloud consists essentially of significant, centralized data sources served by data-center-grade hardware. Considerable research is put into remotely accessed RDF triple stores that can store an ever-increasing amount of triples, most of which are becoming inaccessible during Semantic Web conferences because of saturated Internet connectivity. Citizens living in countries with limited infrastructures experience this daily. Low data center availability, lack of high-end hardware, and unstable infrastructure call for alternative design patterns to data hosting and sharing.

The rapid progress of small-scale devices presents numerous opportunities. They offer low-cost, robust solutions that are easily deployed in contexts where only low electricity power is available. Experience shows that the popular RaspberryPi and the PlugPC from GlobalScale technology can deploy effective ICT solutions (see <http://internet-in-a-box.org>, <http://brck.com>, and www.marvell.com/solutions/education/cloud-computer-for-smile.jsp).

Linked data tools must run on these devices to turn them into low-cost data prosumers, contributing and consuming data in either local or interconnected networks. This implies triple stores and reasoners that perform well on small hardware, and new design paradigms for describing entities without relying on Internet connectivity. In the Entity Registry System (ERS) project (<http://worldwidesemanticweb.org/projects/entity-registries>), we tackle



Figure 1. Example setup for the Entity Registry System. It involves four XO-1 laptops in a mesh network and a RaspberryPi model B data bridge.

the latter issue by providing a system for describing resources that follows the RDF data model, and works on small hardware and under various connectivity contexts. Figure 1 shows a typical setup involving 4 XO-1 laptops locally connected and a RaspberryPi bridging this local network with others.

Multimodal Interfaces

Text-based forms, tables, and desktop-oriented interfaces are currently the primary way of consuming and producing data. Besides the yet-to-be-solved issue that many of the tools produced to deal with RDF data have an entry barrier that is too high for the average user, the bigger issue is that these tools are targeted only to ICT-literate individuals⁸ and cover only one mode of interaction – most often in one language, too.

More effort needs to go into the design of multimodal interfaces for interacting with linked data. One example from which to get inspiration is Ushahidi (www.ushahidi.com/products), an information collection, visualization, and mapping tool that lets users interact with content via SMS messages, and desktop and phone applications. We must also account for basic mobile phones' high penetration rate and the progress made on manufacturing affordable and robust tablets – such as the

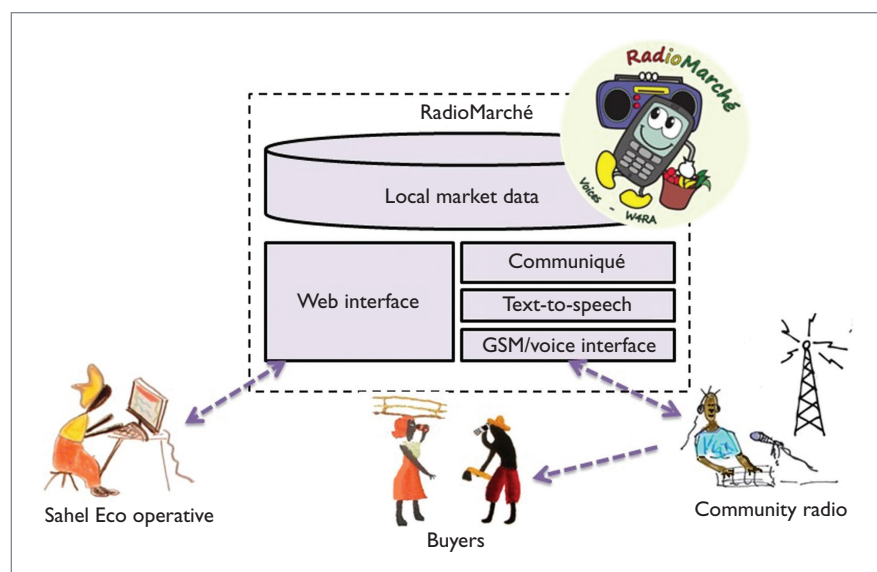


Figure 2. A multimodal interface. Farming goods consumers and producers as well as local radio stations use Radiomarché to share market price data in their own language, using the basic phone and radio they already have.

Akash (<http://akashtablet.com>) – which require new paradigms focused on visual and voice-based direct interaction with linked data.

In fact, the key to providing information sharing solutions that the target population can actually use is implementing the right mix of modalities for interacting with the system. Combining voice-, text-, and icon-based interfaces increases (mobile) ICT interfaces' usability for novice and low-literacy populations.⁹ Illustrating this approach, the RadioMarché system from the EU-funded project Voices (www.mvoices.eu) lets farmers phone a service to announce what they have to sell and name the price they're asking for it.¹⁰ The server then turns the different price information into a voice communiqué that can be broadcast locally via radio or that clients can listen to on demand when calling a specific number (see Figure 2).

Members of the W3C working group Mobile Web for Social Development (MW4D) are paying particular attention to this interplay between mobile phones and ICT (www.w3.org/2008/MW4D/). The work on visual interfaces will present its own set of challenges.

Visual communication depends considerably on the cultural background of its users,¹¹ and such interfaces will have to accommodate various contexts to be usable by the target public.

Contextual Data

Following the open data movement, many governments, non-governmental organizations, and other institutions involved with development are opening and sharing their data, some via linked open data technology. Examples include the International Aid Transparency Initiative (IATI), the World Bank, the Institute of Development Studies (IDS), and the Kenyan Open Data portal (<https://opendata.go.ke/>). The interconnection of development-related datasets is expected to bring insight into how aid money is spent and the actual, observable progress resulting from this aid. So far, partially closed data and isolated sources makes such an audit a daunting task.

As much as open data portals are important, we must also look at individuals' daily data needs and see how linked data could deliver on them. As El Salvador's secretary of state in charge of transparency put it, "When people

do not know what they are going to eat the next day, they have other issues to think about, other than having access to information" (www.youtube.com/watch?v=qOS3juRQXR0). Furthermore, if things aren't done right, it could re-enforce the divide between the elite who have access to the technology and the rest of the population who don't.¹² The risk here is that power falls into the hands of whoever has the most information.¹³

Rather than looking only at large-scale data integration and top-down data dissemination, we must design smaller flexible systems for exchanging bits of locally relevant data, eventually using local semantics that are expressed via local dialects that wouldn't make sense outside a direct user community. Information filtering is also relevant to ensure that the amount of top-down data that a user actually receives isn't overwhelming and is, in fact, relevant. Linked data is a fit solution not only for large-scale information systems but also for these smaller and more focused data flows. We can implement filtering by looking at the description of the circulating resources, and multi-layered and faceted ontologies can be used to distinguish between different levels of interaction (neighbors, community, country, or international).¹⁴ Subsumption reasoning will be a key enabler for this multilevel expressivity, thereby calling for robust reasoners that can work with limited computing resources.

The linked data community has a choice between two approaches for widening the share of the world population that has access to this technology:

- bet on, and contribute to, the global increase of (ICT) capacity and literacy that will, some day, bring all world citizens to the same level with respect to access; or

- act and spend time on adjusting linked data technologies so that those living in underprivileged areas have a higher chance of benefitting from it now.

The community of researchers and open data activists gathered behind the World Wide Semantic Web initiative (see the sidebar) believe in the latter. We are busy changing our best practices, mindsets, and tools while linked data technologies are still young, aiming to make this technology usable by everyone, not just a privileged few. Anyone sharing in this objective is welcome to join us, so don't hesitate! ☐

World Wide Semantic Web

The World Wide Semantic Web is a community of researchers and open data advocates that believe in changing how linked-data-based systems are commonly conceived and deployed (see www.worldwidesemanticweb.org). The goal of affecting a greater share of the world population with these technologies involves working around three axes:

- usage of limited infrastructure (decentralization, limited hardware, limited bandwidth, and so on);
- multimodal and adapted interfaces (voice, icons, touch, culturally localized, and so on); and
- contextualized data sharing (community level, relevant, trusted, and so on).

In contrast with the "big data" trend and the general upscaling of technologies, we coined the term "downscaling" to refer to the work of fitting linked data into weaker infrastructures, providing multimodal interfaces and improving contextualized data sharing.

References

1. N. Shadbolt and K. O'Hara, "Linked Data in Government," *IEEE Internet Computing*, vol. 17, no. 4, 2013, pp. 72–77.
2. S. Hart and T. London, "Developing Native Capability," *Stanford Social Innovation Rev.*, Summer 2005; www.ssireview.org/articles/entry/developing_native_capability.
3. K.F. Becker, *The Informal Economy – Fact Finding Study*, tech. report, Sida, Mar. 2004; www.rrojasdatabank.info/sida.pdf.
4. *The Barahogon Regreening Roadshow*, tech. report, Sahel Eco, 2011; www.saheleco.net/projects/SAHEL_ECO-Regreening_Roadshow_091108r_APERReport.pdf.
5. A. Pentland, R. Fletcher, and A. Hasson, "DakNet: Rethinking Connectivity in Developing Nations," *Computer*, vol. 37, no. 1, 2004, pp. 78–83.
6. L. Raftree, "Open Data, Open Government, and Critical Consciousness," blog, 18 Jan. 2013; <http://lindaraftree.com/2013/01/18/open-data-and-critical-consciousness/>.
7. C. Guéret et al., "Is Data Sharing the Privilege of a Few? Bringing Linked Data to Those Without the Web," *Proc. 10th Int'l Semantic Web Conf.*, 2011, pp. 4–7; <http://wp.me/p1ffiZ-4r>.
8. T. Davies, D. Edwards, and K. Services, "Emerging Implications of Open and Linked Data for Knowledge Sharing in Development," *IDS Bull.*, vol. 43, no. 5, 2013, pp. 117–127.
9. I. Medhi et al., "Designing Mobile Interfaces for Novice and Low-Literacy Users," *ACM Trans. Computer-Human Interaction*, vol. 18, no. 1, 2011, article 2.
10. N. Gyan et al., "Voice-Based Web Access in Rural Africa," *Proc. Web Science Conf.*, 2013; www.few.vu.nl/~vbr240/publications/gyanWebSci2013.pdf.
11. K. Reinecke and A. Bernstein, "Knowing What a User Likes: A Design Science Approach to Interfaces that Automatically Adapt to Culture," *MIS Quarterly*, vol. 37, 2013, pp. 427–453; www.misq.org/skin/frontend/default/misq/pdf/V37I2/ReineckeBernstein.pdf.
12. M. Powell, T. Davies, and K. Taylor, "ICT For or Against Development? An Introduction to the Ongoing Case of Web3," working paper, IKM Emergent, 2012; http://wiki.ikmemergent.net/files/1204-IKM-Working_Paper_16-WEB3-Mar_2012-2.pdf.
13. T. Unwin, *ICT4D: Information and Communication Technology for Development*, Cambridge Univ. Press, 2009.
14. C. Guéret et al., "Genericity versus Expressivity: An Exercise in Semantic Interoperable Research Information Systems for Web Science," Computing Research repository, arXiv preprint 1304.5743, 2013; <http://arxiv.org/abs/1304.5743>.

Christophe Guéret is a research associate at the Data Archiving and Networked Services institute and the eHumanities group of the Royal Netherlands Academy of Arts and Sciences (KNAW). His research activities are

centered around the design of decentralized interconnected knowledge systems and their social and societal implications, especially in the ICT for development (ICT4D) context. Guéret received a PhD at the intersection of computational intelligence and knowledge representation from the University of Tours, France. Contact him at christophe.gueret@dans.knaw.nl or via [@cgueret](https://twitter.com/cgueret).

Victor de Boer is an assistant-professor at Vrije Universiteit Amsterdam. His research focuses on linked data and Semantic Web technologies, specifically, extracting, modeling, and interacting with heterogeneous data in the domain of digital humanities and ICT4D. Contact him via <http://victordeboer.com>.

Stefan Schlobach is an assistant professor at Vrije Universiteit Amsterdam. His current research focuses on using nonstandard techniques and semantics for reasoning and querying, scalable reasoning through approximation and parallelization, and reasoning services for ontology languages such as mapping, explanation, or abduction. Schlobach received a PhD in knowledge extraction in description logics from the University of London. Contact him at k.s.schlobach@vu.nl.

cn Selected CS articles and columns are also available for free at <http://ComputingNow.computer.org>.