



# Royal Netherlands Academy of Arts and Sciences (KNAW) KONINKLIJKE NEDERLANDSE AKADEMIE VAN WETENSCHAPPEN

## The Entity Registry System: Publishing and Consuming Linked Data in Poorly Connected Environments

Guéret, Christophe; Cudré-Mauroux, Philippe

**published in**  
ERCIM News  
2014

**document version**  
Early version, also known as pre-print

[Link to publication in KNAW Research Portal](#)

### **citation for published version (APA)**

Guéret, C., & Cudré-Mauroux, P. (2014). The Entity Registry System: Publishing and Consuming Linked Data in Poorly Connected Environments. *ERCIM News*, (97), 54-55. <http://ercim-news.ercim.eu/en97/ri/the-entity-registry-system-publishing-and-consuming-linked-data-in-poorly-connected-environments>

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the KNAW public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the KNAW public portal.

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**E-mail address:**  
[pure@knav.nl](mailto:pure@knav.nl)

## View other issues

### Contents: ERCIM News 97

# The Entity Registry System: Publishing and Consuming Linked Data in Poorly Connected Environments

by Christophe Guéret and Philippe Cudré-Mauroux

Sixty-five percent of the world's population is deprived of ICT-enhanced data-sharing because of poor access to digital technology. The ExaScale Infolab and Data Archiving and Networked Services (DANS) have designed a new framework, the "Entity Registry System" (ERS), and produced a reference implementation making it possible to publish and consume Linked Open Data in poorly connected environments.

There is no longer any doubt about the social and economic value of accessible, open data. Leveraging the pervasive presence of the Web, Linked Data principles and Web-based portals are promising technologies for implementing data-sharing applications. But the trivial architectural assumptions around data connectivity and availability of large-scale resources keeps ICT-enhanced data sharing out of reach for about 65% of the world's population. The majority of the world's population is digitally disconnected from the rest of the world, and are thus deprived of (Linked) Open Data, and its associated benefits [1].

Work on the Entity Registry System (ERS) started in 2012 as a joint project between the ExaScale Infolab and Data Archiving and Networked Services (DANS) with a generous grant from Verisign Inc. This work is placed in the context of the World Wide Semantic Web initiative aimed at ensuring that the Semantic Web becomes a reality for everyone (see <http://worldwidesemanticweb.org>). Out of the three challenges - infrastructures, interfaces and relevancy - that the World Wide Semantic Web initiative encompasses, ERS tackles the first one.

In contrast to established economies and city dwellers, developing economies and rural populations typically do not have a high speed and stable access to the internet. They also lack access to Web-based data portals where they could publish their new data and consume third-party datasets. The need for sharing open data in such environments is real, however. Monitoring environmental hazards, online medical help, emergency response, and remote education are but a few examples of data sharing scenarios that are essential to the development of emerging countries [1,2]. Technology is increasingly available to such populations, though still in limited forms: solar-powered access to low-speed Internet, local mesh networks, community radio, usb-sticks to carry data, low-cost computers, basic mobile phones, etc. In this context, our goal is to design an approach to publish and consume Linked Open Data using such technologies. More specifically, we decided to focus on existing and relatively cheap equipment such as XO-1 laptops, RaspberryPis, and mesh networks.

In a nutshell, ERS creates global and shared knowledge spaces through series of statements. For instance, "Amsterdam is in the Netherlands" is a statement made about the entity "Amsterdam" relating it to the entity "the Netherlands". Every user of ERS may want to either de-reference an entity (for instance, by asking for all pieces of information about "Amsterdam") or contribute to the content of the shared space by adding new statements. This is made possible via "Contributors" nodes, one of the three types of node defined in our system. Contributors can interact freely with the knowledge base. They are responsible for publishing their own statements but cannot edit third-party statements. Every set of statements about a given entity contributed by a single author is wrapped into a unique named graph to avoid conflicts and enable provenance tracking. In a typical ERS deployment, the end-user machines (eg, the XO-1 laptops) work as Contributor nodes. Two Contributors in a closed P2P network can freely create and share Linked Open Data. In order for them to share data with another closed group of Contributors, we introduce "Bridges". A Bridge is a relay between two closed networks using the internet or any other form of direct connection to share data. Two closed communities, for

example two schools that are willing to share data, can each set up one Bridge and connect these two nodes to each other. The Bridges will then collect and exchange data coming from the Contributors. These bridges are not Contributors themselves, they are just used to ship data (named graphs) around and can be shut-down or replaced without any data-loss. Lastly, the third component we define in our architecture is the "Aggregator". This is a special node every Bridge may push content to. As its name suggests, an Aggregator is used to aggregate entity descriptions that are otherwise scattered among all the Contributors. When deployed, an aggregator can be used to access and expose the global content of the knowledge space or a subset thereof.



*Figure 1: an example of ERS deployment that could be in a school. The four XO laptops are Contributors using the RaspberryPi as a Bridge to asynchronously exchange data with another school.*

An ERS deployment may contain any number of Contributors, Bridges and Aggregators depending on the data-sharing scenario at hand. The system is, however, always "offline by default", relying only on Contributors as persistent data stores, and taking advantage of data flows that are biased towards locally relevant and potentially crowdsourced data. Figure 1 shows a complete test deployment using four Contributors (XO-1 laptop) and a bridge (RaspberryPi model B) connected through a mesh network.

ERS has been developed in Python and Java. It uses CouchDB and CumulusRDF to store the RDF data serialized as JSON-LD. The code of our reference implementation is freely available under an open source license (see <http://ers-devs.github.io>). We are currently working on extending it to support further data sharing scenarios and to make it usable on more hardware.

**Links:**

<http://ers-devs.github.io>

<http://dans.knaw.nl>

<http://exascale.info>

<http://worldwidesemanticweb.org>

**References:**

- [1] F. Estefan, C. Spruill, S. Lee: "Bringing Open Contracting Data Offline", Open Contracting Blog, 2013, [Online], [http://www.open-contracting.org/bringing\\_open\\_contracting\\_data\\_offline](http://www.open-contracting.org/bringing_open_contracting_data_offline)
- [2] T. Unwin: "ICT4D: Information and Communication Technology for Development", Cambridge University Press, 2009, ISBN 9780521712361.

**Please contact:**

Christophe Guéret  
Data Archiving and Networked Services (DANS), The Netherlands  
E-mail: [christophe.gueret@dans.knaw.nl](mailto:christophe.gueret@dans.knaw.nl)