Web of Data Plumbing

Lowering the Barriers to Entry

Jürgen Umbrich[*]
DERI, National University of Ireland, Galway, Ireland

Hugh Glaser[†]
ECS, University of Southampton, UK

Tuukka Hastrup[‡]
University of Jyväskylä, Finnland

Ian C. Millard[§]
ECS, University of Southampton, UK

Michael Hausenblas[¶]
DERI, National University of Ireland, Galway, Ireland

ABSTRACT

Publishing and consuming content on the Web of Data often requires considerable expertise in the underlying technologies, as the expected services to achieve this are either not packaged in a simple and accessible manner, or are simply lacking. In this poster, we address selected issues by briefly introducing the following essential Web of Data services designed to lower the entry-barrier for Web developers: (i) a multi-ping service, (ii) a meta search service, and (iii) a universal discovery service.

1. MOTIVATION

Web of Data applications [1, 2] depend on an infrastructure that supports the publishing and consumption of RDF-based data. Lowering the barrier of entry by the use of facilities—such as our application template[1]–supports the straight-forward publication of data.

Another part of this infrastructure is concerned with the ability for consumers of the Web of Data to find the data they wish to consume. Typical components in this area are *semantic indexers and semantic search engines*[2], collectively referred to as **SISSE** in the following, such as Falcons, PingTheSemanticWeb.com (PTSW), Sindice, SWSE, and Watson. In order to function, SISSE need to be able to find the data that they republish. As well as normal Web crawling, it is helpful if the publishers of data are able to inform the SISSE of new data that had been published, a process that is known as *pinging*.

[*]juergen.umbrich@deri.org
[†]hg@ecs.soton.ac.uk
[‡]Tuukka.Hastrup@iki.fi
[§]icm@ecs.soton.ac.uk
[¶]michael.hausenblas@deri.org

[1]`http://github.com/tuukka/arc2-starter-pack/`
[2]`http://esw.w3.org/topic/TaskForces/CommunityProjects/LinkingOpenData/SemanticWebSearchEngines`

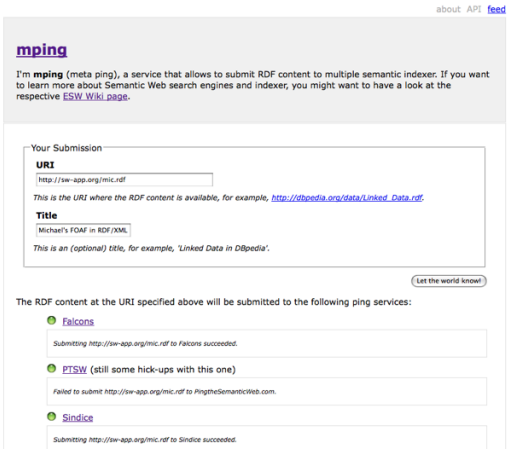

Figure 1: The mping front end.

In this poster, we introduce and briefly discuss three services that simplify the publishing process and support the consumption process, especially concerning the meta discovery issues.

2. META PING

Once RDF content is published, one wants to ensure it can be found. To support the notification process, SISSE usually offer a ping service, enabling the announcement of new content available. We found that two SISSE (PTSW and Sindice) offer XML-RPC-based, pingback APIs[3], whereas three of them (that is, Falcons, SWSE, and Watson) come with proprietary (REST-based) APIs. As it requires considerable web service knowledge and is also laborious to notify all SISSE separately, we have hence developed **mping** (Fig. 1), a *meta ping* service that allows the submission of RDF content to multiple semantic indexers. Additionally, mping offers an Atom news feed of the recent submissions for re-use.

3. META INDEXERS AND SEARCHERS

Another issue—concerning both the publication and the consumption process—is a *meta indexing* facility (compara-

[3]`http://www.hixie.ch/specs/pingback/pingback`

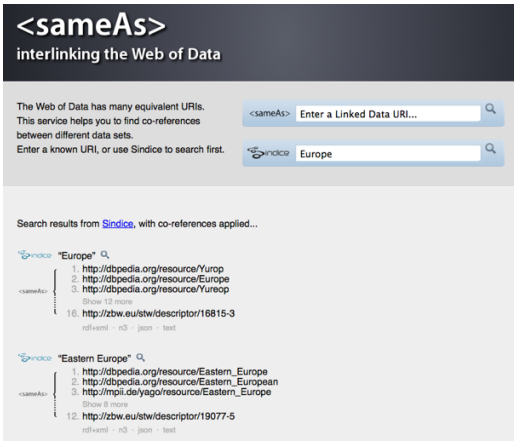

Figure 2: The sameAs front end.

ble to what is available for the Web of Documents[4]) that allows the searching of multiple SISSE and other resources at once. The challenge here (beside accessing the different sources) is how to merge the results. Different ranking methods may be employed by the sources and one has to consolidate and re-rank results. A first step into this direction is our **sameAs** (Fig. 2) service, based on CRS technology[5], that helps Web of Data agents or users to find co-references between different data sets. It does this by providing a consolidation of the linking (indexing) information that is published by indexers and the publishers of the data themselves. A new *meta searching* facility that was announced very recently is **Sigma** [3], a data aggregator released by DERI.

4. META DISCOVERY

From a Web of Data *consumer* perspective, it is essential to find data and learn more about the data found. For example, one might choose to consume only Creative Commons licensed data in the application, or data which is suitable for children, or maybe even—on a rather low level—data which is linked with geolocation data. We hence need metadata about the resources, telling us about what to expect and allowing to assess if a certain data source is relevant for a certain application. We will refer to this process of *figuring out the characteristics of data* in the following as **discovery** [4]. As we have reported in [1], discovery of RDF-based data can take many forms, such as *Link-based Resource Descriptor Discovery* (LRDD)[6], over Follow-Your-Nose[7] to formal descriptions of datasets using the "Vocabulary of Interlinked Datasets" (voiD) [5].

In Fig. 3, **ULDis** (Universal Link DIScovery client) is depicted. With ULDis, we have developed a LRDD-based discovery service that is able to handle the *Extensible Resource Descriptor* (XRD)[8], the *Protocol for Web Description Resources* (POWDER)[9], as well as voiD.

[4] http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/MetaSearch.html
[5] http://eprints.ecs.soton.ac.uk/17587
[6] http://tools.ietf.org/html/draft-hammer-discovery
[7] http://esw.w3.org/topic/FollowLinksForMoreInformation
[8] http://www.hueniverse.com/hueniverse/2009/03/xrd-sneakpeek
[9] http://www.w3.org/TR/powder-dr/

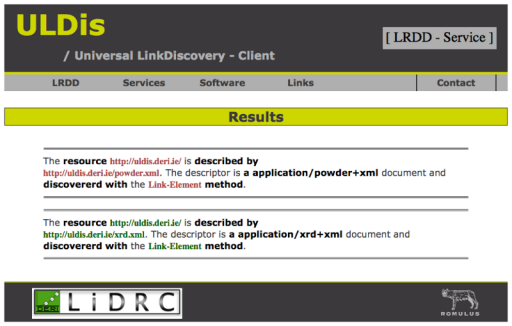

Figure 3: The ULDis front end.

5. DISCUSSION

We have identified and addressed three issues concerning publication and consumption in the Web of Data; it is important to point out that the here presented technologies offer both front-end as well as REST-based interfaces for programatic access. In order to successfully reach out to Web developers, we believe that such simple and appropriate services are required to allow these developers to conveniently and effectively publish and consume RDF-based data. There are certainly other meta services one can think of. One class of services, for example, are meta query, or distributed query systems. As we have pointed out in [5] already, voiD can be and indeed is used for optimising federated queries. We understand, however, that this is subject to further research, holding a number of challenges on it own. We call on the research community at large to start addressing issues such as these, and we ourselves will continue to do so, for example by advancing understanding of Web of Data application development.

The demo for this poster covers (i) the **mping** service[10] in the publication area, (ii) the *sameAs* service[11] as well as **Sigma**[12] in the publication and consumption area, and (iii) **ULDis**[13] in the consumption area for lowering the entry barrier of Web developers concerning Web of Data applications.

6. REFERENCES

[1] M. Hausenblas. Linked Data Applications. Technical Report, DERI, 2009.
[2] M. Hausenblas. Exploiting Linked Data to Build Web Applications. *IEEE Internet Computing*, 13(4):68–73, 2009.
[3] R. Cyganiak, M. Catasta, and G. Tummarello. Towards ECSSE: live Web of Data search and integration. In *WWW09: Semantic Search Workshop*, Madrid, Spain, 2009.
[4] J. Umbrich, M. Hausenblas, P. Archer, E. Hammer-Lahav, and E. Wilde. Discovering Resources on the Web. Technical Report, DERI, 2009.
[5] K. Alexander, R. Cyganiak, M. Hausenblas, and J. Zhao. Describing Linked Datasets. In *WWW09: Linked Data on the Web*, Madrid, Spain, 2009.

[10] http://ld2sd.deri.org/mping/
[11] http://sameas.org/
[12] http://sig.ma/
[13] http://uldis.deri.ie/