<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Web of Data Plumbing - Lowering the Barriers to Entry</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Umbrich, Jürgen; Hastrup, Tuukka; Hausenblas, Michael</td>
</tr>
<tr>
<td><strong>Publication Date</strong></td>
<td>2009</td>
</tr>
<tr>
<td><strong>Publication Information</strong></td>
<td>Jürgen Umbrich, Hugh Glaser, Tuukka Hastrup, Ian C. Millard, Michael Hausenblas &quot;Web of Data Plumbing - Lowering the Barriers to Entry&quot;, Posters and Demos at the 8th International Semantic Web Conference (ISWC 2009), 2009.</td>
</tr>
<tr>
<td><strong>Item record</strong></td>
<td><a href="http://hdl.handle.net/10379/1123">http://hdl.handle.net/10379/1123</a></td>
</tr>
</tbody>
</table>
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, Finland

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 template1—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 engines2, 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
1http://github.com/tuukka/arc2-starter-pack/
2http://esw.w3.org/topic/TaskForces/CommunityProjects/LinkingOpenData/SemanticWebSearchEngines

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 APIs3, 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-

3http://www.hixie.ch/specs/pingback/pingback
Figure 2: The sameAs front end.

able to what is available for the Web of Documents\textsuperscript{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\textsuperscript{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\textsuperscript{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\textsuperscript{4}. As we have reported in [1], discovery of RDF-based data can take many forms, such as Link-based Resource Descriptor Discovery (LRDD)\textsuperscript{6}, over Follow-Your-Nose\textsuperscript{7} to formal descriptions of datasets using the “Vocabulary of Interlinked Datasets” (voiD)\textsuperscript{8}.

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)\textsuperscript{9}, the Protocol for Web Description Resources (POWDER)\textsuperscript{10}, as well as voiD.

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 programmatic 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 its 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\textsuperscript{10} in the publication area, (ii) the sameAs service\textsuperscript{11} as well as Sigma\textsuperscript{12} in the publication and consumption area, and (iii) ULDis\textsuperscript{13} in the consumption area for lowering the entry barrier of Web developers concerning Web of Data applications.

6. REFERENCES


\textsuperscript{4}http://ld2sd.deri.org/mping/
\textsuperscript{10}http://sameas.org/
\textsuperscript{11}http://uldis.deri.ie/