

S2S Architecture and Faceted Browsing Applications

Eric Rozell
Tetherless World Constellation
Rensselaer Polytechnic Institute
Troy, New York, United States
rozele@rpi.edu

Jin Zheng
Tetherless World Constellation
Rensselaer Polytechnic Institute
Troy, New York, United States
zhengj3@rpi.edu

Peter Fox
Tetherless World Constellation
Rensselaer Polytechnic Institute
Troy, New York, United States
pfox@cs.rpi.edu

Jim Hendler
Tetherless World Constellation
Rensselaer Polytechnic Institute
Troy, New York, United States
hendler@cs.rpi.edu

ABSTRACT

This demo paper will discuss a search interface framework designed as part of the Semantic eScience Framework project at the Tetherless World Constellation. The search interface framework, S2S, was designed to facilitate the construction of interactive user interfaces for data catalogs. We use Semantic Web technologies, including an OWL ontology for describing the semantics of data services, as well as the semantics of user interface components. We have applied S2S in three different scenarios: (1) the development of a faceted browse interface integrated with an interactive mapping and visualization tool for biological and chemical oceanographic data, (2) the development of a faceted browser for more than 700,000 open government datasets in over 100 catalogs worldwide, and (3) the development of a user interface for a virtual observatory in the field of solar-terrestrial physics. Throughout this paper, we discuss the architecture of the S2S framework, focusing on its extensibility and reusability, and also review the application scenarios.

Categories and Subject Descriptors

H.5.2 [Information Interfaces and Presentation]: User Interfaces

Keywords

Faceted Search, Semantic Web, Ontologies, Widgets

1. INTRODUCTION

At the Tetherless World Constellation (TWC), we have developed the S2S¹ framework to support the deployment of user interfaces (UIs) for data catalog services. This framework was originally designed in the context of oceanography for the purpose of creating customizable “data dashboards”, which scientists could use as a one-stop shop for their various data needs. Oceanography is a particularly interdis-

¹S2S is an ambiguous acronym, originally standing for “Seafloor to Surface” data dashboard, its scope is now broader.

Copyright is held by the International World Wide Web Conference Committee (IW3C2). Distribution of these papers is limited to classroom use, and personal use by others.

WWW 2012 Companion, April 16–20, 2012, Lyon, France.
ACM 978-1-4503-1230-1/12/04.

ciplinary field, and the scientists pull data from a variety of sources in their day-to-day research activities. The scientists have a tendency to pull data from only sources they are familiar with, to avoid learning curves for new UIs. Search paradigms, such as faceted search, are becoming increasingly popular, and can be used in data dashboards to lower these learning curves [10]. Data managers in the oceanographic community were generally favorable to the idea of a data dashboard platform, as any assistance in the development of sophisticated UIs allows them to focus more resources on improving data management practices and services offered. The framework has evolved beyond this initial concept of data dashboards to meet some of the goals of the Semantic eScience Framework² (SeSF) project, namely for the deployment of virtual observatory portals.

S2S provides a back-end system, the S2S Server, for executing queries over multiple Web service description standards with semantic annotation capabilities (e.g., OpenSearch³ and SAWSDL⁴) and also SPARQL data sources described using the FacetOntology [9]. We use an adapter pattern to enable a uniform query interface across heterogeneous description standards. We have also created an extensible front-end, the S2S User Interface, for deploying faceted and hierarchical search UIs. In addition to meeting the core characteristics of faceted browsers, the front-end architecture also implements an innovative proposal from [7]. In [7], the author discusses the use of suitable “widgets” for each facet in an interface. In particular, based on the kind of data presented as facet values (e.g., nominal, ordinal, or quantitative), some UI mechanisms are more suitable for constraining those facets than others. As an example, a tree map [3] or pie chart may be useful for constraining facets with quantitative values, but would not be suitable for searching facets with nominal values. S2S takes a generic approach to matching facets and results with UI “widgets”.

In this demo paper we discuss the back-end and front-end architectures developed for S2S, and describe projects that have used S2S to deploy UIs for their data catalogs. We end with a discussion of future work and challenges addressed.

²SeSF, <http://tw.rpi.edu/web/project/sesf>

³OpenSearch, <http://www.opensearch.org>

⁴SAWSDL, <http://www.google.com>

2. ARCHITECTURE

The S2S Framework combines a proxy interface to semantically-annotated Web services (the S2S Server) with a portlet⁵-like system for configurable user interfaces (the S2S User Interface). The benefit to combining these two types of systems (generic proxy to Web services and configurable UI) is that such a framework can be used to dynamically generate user interfaces based on data flow; however, not much research has occurred to date for UI to data response matching for S2S.

2.1 S2S Server Architecture

The S2S Server exposes two HTTP services, a metadata service and a data service proxy. Figure 1 provides an overview of this back-end architecture. The metadata service is an endpoint providing information about data services, such as their inputs, outputs, and available operations. The metadata service is also used to provide information about UI widgets, such as entry points for rendering data service responses. Data services and UI widgets are described using an RDF vocabulary, shown in Figure 2. The proxy service uses an adapter pattern to “ground” semantic requests to the S2S Server into specific requests required for individual Web services. Adapters in S2S achieve two purposes, avoiding the need for redundant annotations, where annotation might be implicit in the syntactic description of a Web service, and enabling a single protocol to access various Web services and standards. Each adapter is responsible for parsing the syntactic description of the Web service, and retrieving RDF metadata for the service when necessary. This adapter pattern provides extensibility to the framework. For instance, we have currently implemented adapters for OpenSearch-described services and SPARQL services described using the mSpace FacetOntology [9]; we are also developing adapters for SAWSDL-described services, to support a broader variety of the use cases and existing services. Although primarily used to collect details for the invocation of Web services, the metadata service in the architecture could also be employed to facilitate broader discovery and re-use through the use of other Web service ontologies, such as OWL-S [4].

2.2 S2S User Interface Architecture

While S2S is intended to support a variety of UIs to data services, including workflows, the current front-end has been designed to implement faceted and hierarchical search. The front-end is based on an abstract model for UI components referred to as “widgets”. Widgets are used to render information in a browser, and to facilitate user input, both through form inputs, and visualizing contextual information. In the context of faceted browse, the interface supports two types of widgets, facet widgets and result widgets. Facet widgets display facet values and allow the user to execute actions typically supported by faceted browsers, such as zoom and pivot [10]. Result widgets display the results of the search, such as datasets from a catalog. Widgets in S2S are analogous to portlets; their primary difference is that widgets are described using RDF (as opposed to XML), and that the procedural code for the widgets is potentially Web-addressable (e.g., JavaScript) as opposed to being encapsulated in a portlet server. The RDF metadata associated

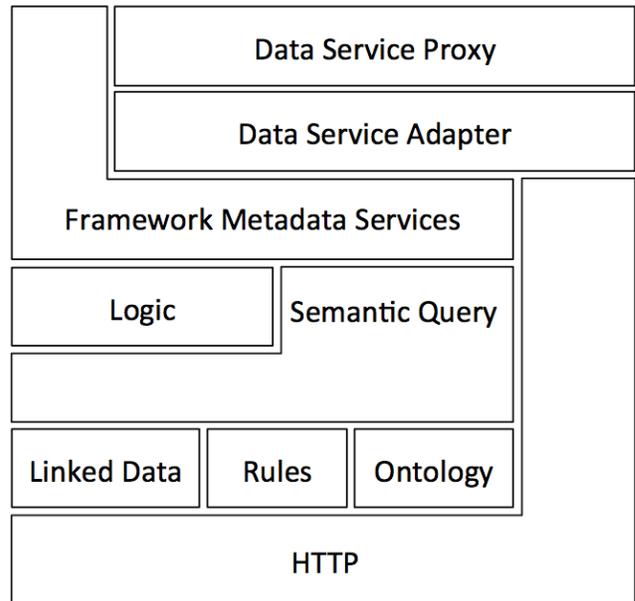


Figure 1: The S2S back-end stack.

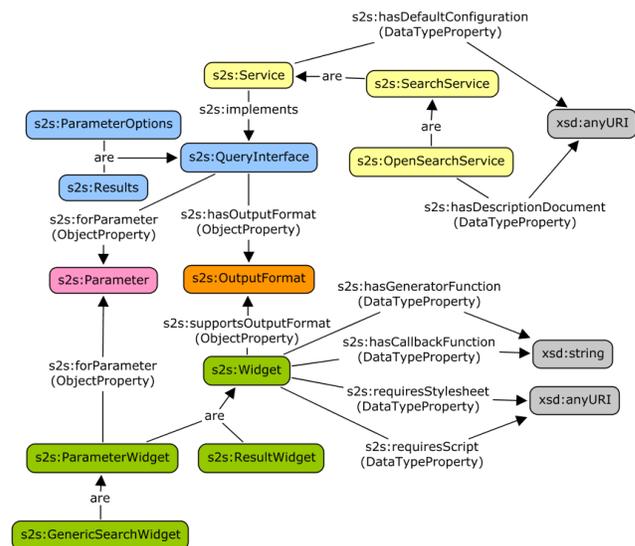


Figure 2: A compressed version of the S2S vocabulary used to describe data services.

⁵An example being the Java Portlet Specification 2.0

with each widget describes (1) the type of data inputs supported by the widget, (2) the procedural component used to generate and update the widget, such as the location of JavaScript files, and (3) the function names of specific entry points for the widget. The S2S front-end uses the metadata to match widgets with data service responses, dynamically load Web resources needed to render the widget, and call specific entry points when the widget is generated or asynchronous data requests complete, respectively. The metadata can also specify if a facet widget is generic, meaning it can be used for any facet, or if it is intended to be used with a specific facet (e.g., a map widget for constraining a geographic facet). Figure 2 also contains the vocabulary terms and relationships for S2S widgets.

3. APPLICATION SCENARIOS

We implemented faceted browsers and other search interfaces for multiple data sources. Here we discuss three applications, which cover: (1) embedding search capabilities from S2S into existing UIs for data access and analysis, (2) embedding the S2S search into a content management system (CMS), and (3) using S2S as a portal for a virtual observatory framework. Scenario 1 has been demonstrated with the deployment of S2S for the Biological and Chemical Oceanographic Data Management Office⁶ (BCO-DMO). The implementation of the International Open Government Dataset Search (IOGDS) by the Linking Open Government Data⁷ group at TWC demonstrates scenario 2. Lastly, we are working towards a generic portal solution for virtual observatories based on the progress in the SeSF project at TWC, which we use to demonstrate scenario 3.

BCO-DMO is a data catalog provider for oceanographic datasets in the domains of biology and chemistry from Woods Hole Oceanographic Institution. We developed an OpenSearch Web service to access their data catalog. In addition to the basic search parameters defined by OpenSearch and its extensions, we have created custom parameters that suit the use cases BCO-DMO has established for faceted browse. These parameters include the projects and programs for which the dataset was collected, the funding number that supported the data collection, as well as the instruments used to collect the datasets. We have embedded a faceted search panel in their current implementation of the open source MapServer⁸. Figure 3 provides a screenshot of the combined S2S and MapServer implementation.

IOGDS[1] provides a faceted search over more than 700,000 datasets from over one hundred data catalogs representing more than 30 countries and international organizations. We have developed an OpenSearch service to transform basic HTTP requests into SPARQL [8] queries over the linked data catalog created from integrating international dataset catalogs with a common vocabulary. We have created custom parameters for zooming into countries, catalogs, agencies, and keywords used to describe the datasets. The IOGDS was originally going to be implemented in Exhibit [2], however the version of Exhibit available at that time was not able to scale the number of datasets provided by IOGDS. This was a major strength for S2S at the time it was developed. The faceted search interface has been embedded in

a Drupal CMS. There is also a live demo⁹ available on the Web.

The last scenario is currently being developed for the SeSF project. SeSF aims to develop tools and methodologies based on Semantic Web technologies for the design and construction of virtual observatories and scientific data portals [5]. One of the initiatives is to design an ontology framework that can be extended and mapped to by individual projects and data providers. By using these ontologies, the projects will be able to use the tools being developed under SeSF. We are building a generic OpenSearch service, which can be used to search any RDF data store that uses, or maps to, terms from the SeSF ontology framework. Currently, within SeSF, we have implemented a prototype UI to the Virtual Solar Terrestrial Observatory (VSTO). The VSTO prototype includes both hierarchical and temporal facets to upper atmospheric and solar terrestrial metadata, with more than 100,000,000 triples covering the entire metadata collection.

4. DISCUSSION

There are a number of developments in the pipeline for S2S. Some mentioned earlier include expanding the Web service ontology and metadata services to offer discovery-level functionality based on OWL-S service profiles. Another area we are interested in, related to data and Web service discovery, is federated search. Using the proxy architecture of S2S, we can build a federated data discovery layer that focuses on parsing search parameters from free-text or semi-structured queries, and forwards those search parameters to the uniform interface provided by the S2S server. Lastly, we are interested in expanding the vocabularies and architecture to support a workflow system using the widget architecture. One option for this is to integrate the vocabulary and architecture into the Semantic Application Framework (SAF), described in [6]. SAF has a similar philosophy in terms of describing the metadata about inputs and outputs of UI components, but is lacking in areas covered by S2S, such as generic descriptions of Web services and access to data via generic queries to the proxy architecture.

We have presented S2S, a UI framework for data catalog services, and have briefly described some of the applications that we have deployed. S2S has taken a Web service approach to faceted browse interfaces, shifting some of the difficult computation for managing facets and results to server-side implementations. However, in doing so, S2S has enabled a greater degree of scalability, most notably for our IOGDS portal, which searches over more than 700,000 datasets. It has also been facilitative of the widget architecture, which implements some of the novel concepts described in [7] for matching user interface mechanisms with facets based on classes of facet values. Links to live demonstrations and videos are available on the Web¹⁰.

5. ACKNOWLEDGMENTS

This work was started during a Summer Student Fellowship at Woods Hole Oceanographic Institution, funded by the Academic Programs Office. Continued research and development on the S2S framework has been supported by

⁶BCO-DMO, <http://www.bco-dmo.org>

⁷LOGD, <http://logd.tw.rpi.edu>

⁸MapServer, <http://www.mapserver.org/>

⁹IOGDS, http://logd.tw.rpi.edu/demo/international_dataset_catalog_search_v_12

¹⁰Live demos and videos, <http://tw.rpi.edu/web/project/sesf/workinggroups/s2s/www-demos>

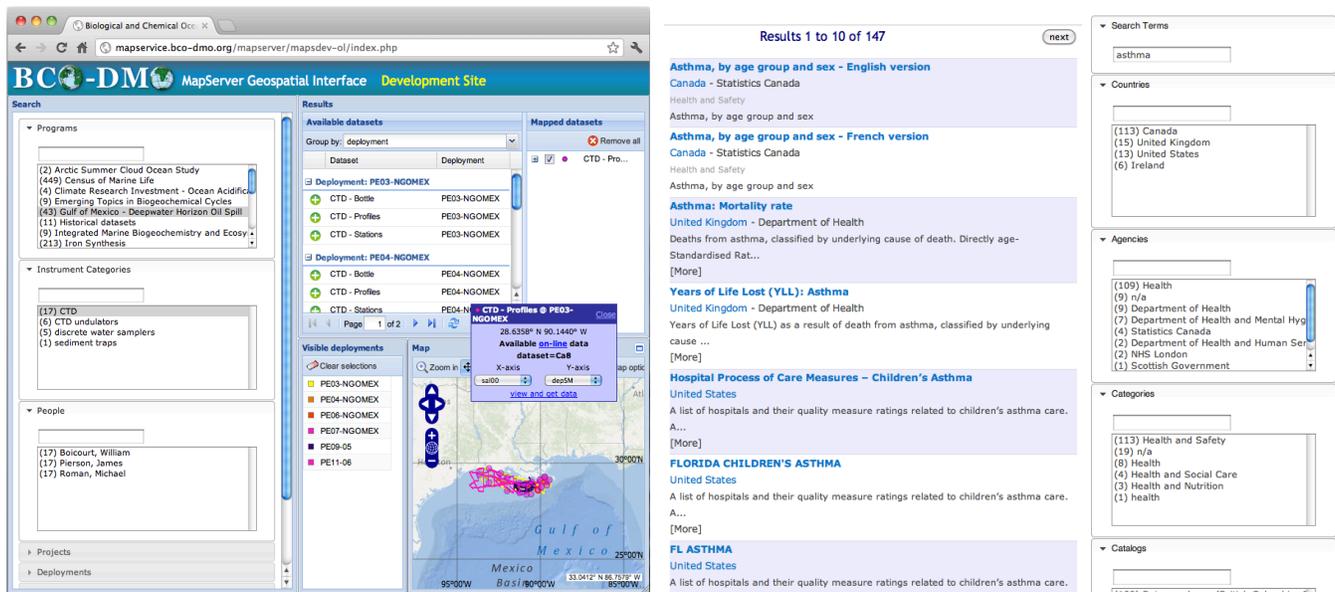


Figure 3: Screenshots of the BCO-DMO search interface (left) and the IOGDS search interface (right). S2S allows a significant amount of customization, whether the interface will be embedded in a CMS such as Drupal (as is the case for IOGDS) or if the interface will push results to a custom application (such as the MapServer for BCO-DMO). The facets are configurable by both the UI provider and the end-user. S2S provides flexibility in the kinds of use cases that can be implemented. Whether it is finding data related to the Deep Water Horizon Oil Spill (left) or related to asthma rates in different countries (right), S2S provides the framework for rapidly deploying faceted browsers and other search interfaces.

the Semantic eScience Framework project in the Tetherless World Constellation at Rensselaer Polytechnic Institute, funded by NSF Office of Cyberinfrastructure award number OCI-0943761. Partial funding for this work has been provided by a gift from Microsoft Research to the Tetherless World Constellation at RPI.

6. REFERENCES

- [1] J. Erickson, Y. Shi, L. Ding, E. Rozell, J. Zheng, and J. Hendler. Twc international open government dataset catalog. In *Proceeding of the Seventh International Conference on Semantic Systems*, 2011.
- [2] D. F. Huynh, D. R. Karger, and R. C. Miller. Exhibit: lightweight structured data publishing. In *Proceedings of the 16th international conference on World Wide Web, WWW '07*, pages 737–746, New York, NY, USA, 2007. ACM.
- [3] B. Johnson and B. Shneiderman. Tree-maps: a space-filling approach to the visualization of hierarchical information structures. In *Proceedings of IEEE Conference on Visualization*, pages 284–291, October 1991.
- [4] D. Martin, M. Paolucci, S. McIlraith, M. Burstein, D. McDermott, D. McGuinness, B. Parsia, T. Payne, M. Sabou, M. Solanki, N. Srinivasan, and K. Sycara. Bringing semantics to web services: The owl-s approach. In J. Cardoso and A. Sheth, editors, *Semantic Web Services and Web Process Composition*, volume 3387 of *Lecture Notes in Computer Science*, pages 26–42. Springer Berlin / Heidelberg, 2005.
- [5] D. McGuinness, P. Fox, P. West, E. Rozell, S. Zednik, and C. Chang. Progress toward a semantic escience framework; building on advanced cyberinfrastructure, 2010. AGU Fall Meeting Abstracts.
- [6] E. Patton, D. Difranzo, and D. McGuinness. Saf: A provenance-tracking framework for interoperable semantic applications. In D. McGuinness, J. Michaelis, and L. Moreau, editors, *Provenance and Annotation of Data and Processes*, volume 6378 of *Lecture Notes in Computer Science*, pages 73–77. Springer Berlin / Heidelberg, 2010.
- [7] J. Polowinski. Widgets for faceted browsing. In M. Smith and G. Salvendy, editors, *Human Interface and the Management of Information. Designing Information Environments*, volume 5617 of *Lecture Notes in Computer Science*, pages 601–610. Springer Berlin / Heidelberg, 2009.
- [8] E. Prud'hommeaux and A. Seaborne. Sparql query language for rdf, 2008. W3C Recommendation 15 January 2008.
- [9] D. Smith and m. Schraefel. Interactively using semantic web knowledge: Creating scalable abstractions with facetontology. Technical report, School of Electronics and Computer Science, University of Southampton, 2008.
- [10] M. Stefaner, S. Ferré, S. Perugini, J. Koren, and Y. Zhang. User interface design. In G. M. Sacco, Y. Tzitzikas, and W. B. Croft, editors, *Dynamic Taxonomies and Faceted Search*, volume 25 of *The Information Retrieval Series*, pages 75–112. Springer Berlin Heidelberg, 2009.