

Linked Data Spaces & Data Portability

Kingsley Idehen
OpenLink Software,
10 Mall Road,
Burlington, MA 01803, USA

kidehen@openlinksw.com

Orri Erling
OpenLink Software,
10 Mall Road,
Burlington, MA 01803, USA
oerling@openlinksw.com

ABSTRACT

In the year 2007, the size of the Linked Data injected into the Web grew to several billion RDF triples, served by a network of interlinked data sources that cover domains such as general knowledge, geographic information, people, companies, online communities, films, music, books and scientific publications. Unfortunately, the growth rate of User Generated content from a variety of Web based unstructured and semi-structured data-silos continues to exceed that of structured Linked Data. Thus, we have a pressing need for technology, capable of bridging this broadening divide via transparent generation of Linked Data from existing data-silos on the Web. Our Linked Data technology demonstration explores the use of the OpenLink Data Spaces platform as a solution to this problem.

Categories and Subject Descriptors

H.3.2 [Information Storage]

H.3.3 [Information Search & Retrieval]

General Terms

Management, Performance, Design, Standardization, Languages, Theory

Keywords

Linked Data, Semantic Web, SPARQL, Data Integration, Data Spaces

1. INTRODUCTION

User generated content is growing at an exponential rate behind corporate firewalls and across the Internet in general. The use of Web technologies has been the prime accelerator of the aforementioned growth due to the pervasiveness of Web based distributed collaborative applications. Examples include: Social Networking, Weblogs, Wikis, Shared Bookmark Managers, Photo Sharing, Polls Management, Calendars, Discussion Forums, File Sharing, and Feed Aggregation, to name a few.

The exponential growth of user-generated content has resulted in the growth of silos comprised of unstructured and/or semi-structured content. Unfortunately, these silos have accelerated, rather than decelerated, the imminence of an “information overload” quagmire.

To alleviate the imminent challenges of global information overload, we need to unobtrusively construct a Web of interlinked structured data from today’s data silos comprised of the following:

- RDF based structured data
- Standardized data serialization formats
- HTTP based Unique Identifiers for all Data Items (web resources and abstract & concrete things)
- HTTP based Data Set containers (Data Spaces)
- Data Servers that provide data management and data access services for one or more Data Spaces
- Key infrastructure oriented shared ontologies
- Query Language for interacting with structured data

We identify the items above, collectively, as critical components of Linked Data Spaces: points of presence on the Web that expose structured data via HTTP based URIs.

During this demonstration / presentation session we are going to explore the creation of “Data Junction Boxes in the Clouds” via OpenLink Data Spaces that exploits in-built RDFization Middleware, plus the ability to mesh User Identity and User Data, en route to surmounting the issues and challenges associated with Data Portability attainment.

2. Issues & Challenges

2.1 Data Portability

It’s no secret that data wants to be free of the tyranny of application logic confinement. In recent times, the realization that meshing Identity and Data ownership on the Web are critical requirements of this pursuit of freedom, has resulted in the emergence of a movement for Data Portability as yet another enclave within the broader Open Data movement.

Data portability addresses to key issues: data mobility and data referencing. Today, data mobility through the use of standard data formats for moving data across silos (import and export style) have emerged as the focal point of attention with regards to addressing the proliferation of data silos on the Web. Examples include: RSS 1.0, RSS 2.0, Atom, OPML, FOAF, SIOC, and others. Unfortunately, the ability to reference and de-reference data across data-silos is yet to catch the attention of those pursuing data portability.

2.2 RDFization Middleware

The traditional resistance to RDF adoption, which is critical to Linked Data comprehension and production, comes from the grounding of the RDF Data Model in Graph Theory and the unwillingness of most Web Application developers to interact with data formally. This reality has led to a genre of middleware tools collectively known as RDFizers, that generate RDF on the fly.

With regards, to Linked Data, generating RDF on-the-fly is only part of the equation; the generated RDF must retain the core principles of linked data by providing URIs for physical web accessible resources, concrete entities, and abstract things. Of course, this process must include intelligent production of instance data associated with relevant shared schemas or ontologies.

2.3 Data Junction Boxes in the Clouds

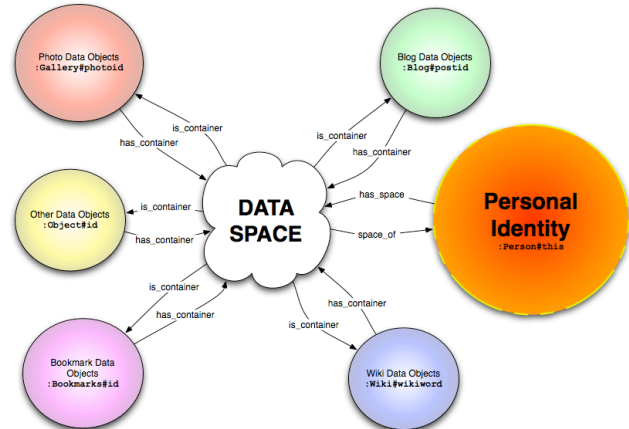
It is our belief that the Linked Data Web will be more distributed than centralized in architecture. We envisage a Linked Data Web comprised of hubs that range in size from large (e.g. DBpedia, Geonames, Zitgist etc.), medium sized group (e.g. RDFized Weblogs, Wikis, Bulletin Boards etc.), and smaller personal hubs enabled by operating system virtualization technologies like Amazon EC2. The medium and smaller hubs are best described as data junction boxes because they act as conduits between existing systems and Linked Data aware User Agents.

This demonstration will demonstrate a Data Space initialization process for end-users that covers:

- Domain Name Registration (e.g. .Name acquisition)
- DNS configuration
- Bonding with existing Web 2.0 platforms Facebook, phpBB3, MediaWiki, Wordpress, Drupal, Del.icio.us, Flickr, and Bugzilla
- Production of a dereferencable URIs that exposed the resulting Data Graph

- Interaction with the resulting data graph via a number of Linked Data aware User Agents

3. Identity & Data Meshing via Linked Data Spaces



4. Links

- http://en.wikipedia.org/wiki/OpenLink_Data_Spaces - OpenLink Data Spaces
- http://en.wikipedia.org/wiki/Virtuoso_Universals_Server - Virtuoso
- <http://myopenlink.net/ods/index.html> - Live OpenLink Data Spaces Demonstration