



LarKC

*The Large Knowledge Collider:
a platform for large scale integrated reasoning and Web-search*

FP7 – 215535

D9.2.v2 Report on Market Observation and Standard Assessment, Update

Coordinator: Frank van Harmelen
(with contributions from many consortium members)

Document Identifier:	LarKC/2008/D9.2.v2
Class Deliverable:	LarKC EU-IST-2008-215535
Version:	1.1
Date:	22-01-2010
State:	Final
Distribution:	Public



EXECUTIVE SUMMARY

This brief report is an update of deliverable D9.2.1-V1. It contains updates since that deliverable concerning the assessment of relevant standardisation activities, and concerning observations of relevant markets. LarKC's general approach to monitoring and participating in standards development remains unchanged from V1 of this deliverable. This report provides observations about recent developments, backup up with numerous pointers to websites that illustrate the developments. Each section briefly mentions to the relevance for LarKC. In the standards area, the following W3C developments are directly relevant to LarKC: SPARQL, OWL2, RIF and RDF/a. These are all well aligned with LarKC's technology. In the marketplace, the most dominant and sustained developments have been those of Linked Open Data (LOD) in general, and its use in government applications (data.gov and data.gov.uk) in particular. The rise of LOD as a dominant model for the Semantic Web is well aligned with LarKC's goals and technology; the lack of business models is a threat to the success of LOD (and LarKC) in the medium term.



DOCUMENT INFORMATION

IST Project Number	FP7 - 215535	Acronym	LarKC
Full Title	The Large Knowledge Collider: a platform for large scale integrated reasoning and Web-search		
Project URL	http://www.larkc.eu/		
Document URL			
EU Project Officer	Stefano Bertolo		

Deliverable	Number	D9.2.1-V2	Title	Report on Market Observation and Standard Assessment, Update
Work Package	Number	WP9	Title	Exploitation and Standardization

Date of Delivery	Contractual	M 18	Actual	M22
Status	version 1.2		final	<input checked="" type="checkbox"/>
Nature	prototype <input type="checkbox"/> report <input checked="" type="checkbox"/> dissemination <input type="checkbox"/>			
Dissemination level	public <input checked="" type="checkbox"/> consortium <input type="checkbox"/>			











Authors (Partner)	VUA			
Responsible Author	Name	Frank van Harmelen	E-mail	Frank.van.Harmelen@cs.vu.nl
	Partner	VU	Phone	+31205987731

Abstract (for dissemination)	In the standards area, the following W3C developments are directly relevant to LarKC: SPARQL, OWL2, RIF and RDF/a. In the marketplace, the most dominant and sustained developments have been those of Linked Open Data (LOD) in general, and its use in government applications (data.gov and data.gov.uk) in particular. The rise of LOD as a dominant model for the Semantic Web is well aligned with LarKC's goals and technology; the lack of business models is a threat to the success of LOD (and LarKC) in the medium term.
Keywords	Standards, market developments

Version Log			
Issue Date	Rev. No.	Author	Change
17 Jan 2010	1.0	Frank van Harmelen	Initial distributed version
21 Jan 2010	1.1	Frank van Harmelen	Expanded with explanatory text
22 Jan 2010	1.2	Frank van Harmelen	Expanded with conclusions



PROJECT CONSORTIUM INFORMATION

Participant's name	Partner	Contact
Semantic Technology Institute Innsbruck, Universitaet Innsbruck	 	Dieter Fensel, Semantic Technology Institute (STI), Universitaet Innsbruck, Innsbruck, Austria, E-mail: dieter.fensel@sti-innsbruck.at
AstraZeneca AB		Bosse Andersson, AstraZeneca Lund, Sweden Email: bo.h.andersson@astrazeneca.com
CEFRIEL - SOCIETA CONSORTILE A RESPONSABILITA LIMITATA		Emanuele Della Valle, CEFRIEL - SOCIETA CONSORTILE A RESPONSABILITA LIMITATA, Milano, Italy, Email: emanuele.dellavalle@cefriel.it
CYCORP, RAZISKOVANJE IN EKSPERIMENTALNI RAZVOJ D.O.O.		Michael Witbrock, CYCORP, RAZISKOVANJE IN EKSPERIMENTALNI RAZVOJ D.O.O., Ljubljana, Slovenia, Email: witbrock@cyc.com
Höchstleistungsrechenzentrum, Universitaet Stuttgart		Georgina Gallizo, Höchstleistungsrechenzentrum, Universitaet Stuttgart, Stuttgart, Germany, Email: gallizo@hlrs.de
MAX-PLANCK GESELLSCHAFT ZUR FOERDERUNG DER WISSENSCHAFTEN E.V.		Lael Schooler, Max-Planck-Institut für Bildungsforschung Berlin, Germany Email: schooler@mpib-berlin.mpg.de
Ontotext AD		Atanas Kiryakov, Ontotext Lab, Sofia, Bulgaria Email: naso@ontotext.com
SALTLUX INC.		Kono Kim, SALTLUX INC, Seoul, Korea, Email: kono@saltlux.com
SIEMENS AKTIENGESELLSCHAFT		Volker Tresp, SIEMENS AKTIENGESELLSCHAFT, Muenchen, Germany, E-mail: volker.tresp@siemens.com
THE UNIVERSITY OF SHEFFIELD		Hamish Cunningham, THE UNIVERSITY OF SHEFFIELD Sheffield, UK, Email: h.cunningham@dcs.shef.ac.uk



<p>VRIJE UNIVERSITEIT AMSTERDAM</p>		<p>Frank van Harmelen, VRIJE UNIVERSITEIT AMSTERDAM, Amsterdam, Netherlands, Email: Frank.van.Harmelen@cs.vu.nl</p>
<p>THE INTERNATIONAL WIC INSTITUTE, BEIJING UNIVERSITY OF TECHNOLOGY</p>		<p>Ning Zhong, THE INTERNATIONAL WIC INSTITUTE, Mabeshi, Japan, Email: zhong@maebashi-it.ac.jp</p>
<p>INTERNATIONAL AGENCY FOR RESEARCH ON CANCER</p>	 <p>International Agency for Research on Cancer Centre International de Recherche sur le Cancer</p>	<p>Paul Brennan, INTERNATIONAL AGENCY FOR RESEARCH ON CANCER, Lyon, France, Email: brennan@iarc.fr</p>
<p>INFORMATION RETRIEVAL FACILITY</p>		<p>John Tait, INFORMATION RETRIEVAL FACILITY Vienna, Austria Email : john.tait@ir-facility.org</p>



TABLE OF CONTENTS

1. STANDARDS ASSESSMENT	7
1.1. SPARQL	7
1.2. RIF	7
1.3. OWL2	8
1.4. RDF/A	8
1.5. OPENLR	9
2. MARKET OBSERVATIONS.....	9
2.1. LINKED OPEN DATA.....	9
2.2. USAGE OF LOD	9
2.3. DEPLOYMENT OF GOODRELATIONS	11
2.4. DATA.GOV AND DATA.GOV.UK.....	10
2.5. LINKED OPEN DRUG DATA	10
2.6. WOLFRAM ALPHA	11
2.7. NEW YORK TIMES TAGS.....	11
2.8. LIBRARY OF CONGRESS NAME-AUTHORITY FILE.....	11
2.9. FREEBASE.....	12
2.10. MICROSOFT PATENT.....	12
2.11. LINEDGEODATA.ORG	12
2.12. COMPANIES HOSTING SEMANTIC WEB DATA	12
2.13. BUSINESS MODELS FOR SEMANTIC TECHNOLOGIES.....	13
2.14. SEARCH ENGINES	13
2.15. DRUPAL SUPPORTING RDF/A	13
2.16. TRIPLE STORES	14
3. CONCLUSION	14



1. Introduction

Members of the consortium are continuously using a variety of monitoring and networking tools to maintain an up to date view of both standardisation activities and market developments. These are (among others): active participation in the blogosphere (both reading and writing), alert-services on major search engines (e.g. “Google alerts”), and use of Web2.0 tooling such as Twitter and Facebook.

This report bundles the observations of consortium members since the first release of this document (D9.2.1, V1). Many of the points have already appeared online, either in the LarKC blog (at <http://blog.larkc.eu>), or on Web2.0 platforms.

2. Standards assessment

LarKC's general approach to monitoring and participating in standards development remains unchanged from V1 of this deliverable. The focus is on MPI, OASIS, Open Grid Forum, and W3C as the main standards bodies, since between them they cover the technologies that LarKC is concerned with (see table 13 of V1 of this deliverable). The LarKC consortium has an active presence in all of these bodies (again see V1 of this deliverable for details).

We list the most relevant standardisation developments since October 2008, and their relevance to LarKC.

2.1. SPARQL

An update for the SPARQL definition has been initiated, with first working drafts produced in October 2009 http://www.w3.org/2009/sparql/wiki/Main_Page. Some of the main innovations in these working drafts are:

- negation and aggregates in queries
- updates to RDF graphs
- a SPARQL protocol
- a vocabulary and discovery mechanism for SPARQL endpoints
- querying under entailment regimes

More recently, a proposal has been made to introduce unique URI's as names for OWL2 profiles. This is in line with comments provided by LarKC consortium members to the OWL2 final call, where we asked (without success) for exactly such URI's to be introduced.

Relevance to LarKC: Many of the proposed revisions and extensions are relevant (and even crucial to LarKC):

- Description and discovery mechanisms for SPARQL endpoints are relevant for IDENTIFY plugins
- A SPARQL protocol is relevant for all plugin APIs
- A need for updating the RDF graphs has been felt in all of LarKC's use-cases
- Querying under entailment regimes is directly relevant for LarKC data-layer.

The architecture interfaces and implementations developed in WP5 should take note of the new SPARQL developments and make sure they are implemented. LarKC could even consider to serve as a reference implementation for some of the items above when the SPARQL recommendation process proceeds.

2.2. RIF

Standardisation of a Rule Interchange Format (RIF) has now reached W3C Candidate Recommendation status.

http://www.w3.org/2005/rules/wiki/RIF_Working_Group,

and

http://www.w3.org/blog/SW/2009/10/02/rif_is_a_w3c_candidate_recommendation.



The Working Group published six Candidate Recommendations. Together, they allow systems using a variety of rule languages and rule-based technologies to interoperate with each other and with Semantic Web technologies. Three of the drafts define XML formats with formal semantics for storing and transmitting rules:

- The RIF Production Rule Dialect (PRD) for rules used in Business Rule Management systems.
- The RIF Basic Logic Dialect (BLD) as a foundation for Logic Programming, classical logic, and related formalisms.
- The RIF Core Dialect as the common subset of PRD and BLD,

Relevance for LarKC: This standard is of importance for reasoner plugins, although not all reasoners are expected to be RIF compliant (since they inhabit spaces simply not covered by RIF). The IRIS reasoner (already ported as one of LarKC's reasoning plugins) is currently aiming to support the BLD dialect.

A major limitation on RIF from LarKC's point of view is its choice for an XML-only encoding of rules. All LarKC's interfaces are designed for exchange of RDF-encoded information. The choice for XML sits uneasily with the view of "rules-as-data". Unfortunately, this is a basic design choice in RIF and unlikely to be revisited.

2.3. OWL2

The OWL2 extension to OWL has become a W3C recommendation. <http://www.w3.org/TR/2009/REC-owl2-overview-20091027/>

Some of the major innovations of OWL2 are (<http://www.w3.org/TR/2009/REC-owl2-new-features-20091027/>):

- Language profiles
- Additional expressivity
- Non-RDF serialisations
- Extended data-type capabilities
- Some useful syntactic sugar constructs
- Simple meta-modelling through punning

Relevance for LarKC: Although much of the additional expressivity defined for OWL2 is not immediately relevant to the LarKC use-cases, some of the lower expressivity profiles defined by OWL2 are relevant to LarKC. Indeed, OWL2 RL has now been adopted as the recommended representation language to be supported by LarKC plugins, replacing the independently developed L2 dialect. L2 had been developed based on an analysis of use-case requirements and on the support-levels in the LarKC data-layer, and turned out to be very close to OWL2 RL. Given LarKC's choice to define the plugin-interface around RDF exchange, the non-RDF serialisation is not immediately relevant to LarKC.

2.4. RDF/a

RDF/a became a W3C recommendation <http://www.w3.org/TR/2008/NOTE-xhtml-rdfa-primer-20081014/>. RDF/a provides a set of XHTML attributes which allow the embedding of RDF statements in regular XHTML browser accessible pages. It allows to turn existing human-visible text and links into machine-readable data without repeating content.

Relevance for LarKC: RDF/a is rapidly becoming a major source of semantic data on the Web, and is hence of importance to IDENTIFY and TRANSFORM plugins, that are concerned with finding and obtaining data-sources.



2.5. OpenLR

Although not covered by our list of standardisation bodies (and in fact not really a "standardisation" at all), the activities at around OpenLR <http://www.tomtom.com/page/openLR> are noteworthy. OpenLR is an initiative lead by TomTom to enable reliable data exchange and cross-referencing using digital maps of different vendors and versions.

Relevance for LarKC: This is potentially relevant to the Urban use-case in WP6.

3. Market Observations

The market for semantic technologies remains a very dynamic place with commercial and technological developments in rapid succession. The list in this section is not an exhaustive survey, but is a "characterisation by snapshots" of the relevant developments in the market place. They have been loosely grouped based on technological and commercial sector.

3.1. Linked Open Data

The most prominent events have with doubt been the developments around Linked Open Data (LOD, <http://www.linkedopendata.org>). The dataset continues to be growing at a rapid pace, is currently estimated at 13 billion triples, having tripled in size in 9 months (march to November 2009). DBPedia remains LOD's central hub. DBPedia's most recent 3.4 release references more than 2.9 million "things" - including (at least) 282,000 persons, 339,000 places, 88,000 music albums, 44,000 films, 15,000 video games, 119,000 organizations, 130,000 species and 4400 diseases.

Relevance for LarKC: LOD is rapidly becoming the main source of semantic data on the Web. LarKC must ensure that both its technical interface and the supported expressivity is compatible with LOD developments. LarKC seems well placed in this space, with its adoption of RDF, SPARQL and OWL2 RL.

3.2. Usage of LOD

Evidence of the usefulness of LOD is beginning to emerge from LOD-usage by external companies who do not have a primary mission in semantic technologies, but who apparently see LOD benefitting their primary mission. Well known examples are Reuter's OpenCalais service, who started to link to public LOD identifiers (instead of using only their own identifiers)

http://www.readwriteweb.com/archives/calais_4_linked_data.php. Calais 4.0 now allows content publishers to integrate their content with Linked Data assets from Wikipedia, GeoNames, the Internet Movie Database (IMDB), Shopping.com and others.

A second, related example is that announcement that local UK newspapers will deploy the Calais service to annotate and link their content <http://3.ly/ZNE>.

Similarly, OpenCalais functionality is being closely integrated into the mainstream Oracle Database product <http://3.ly/e8P>. Oracle users can now extract and link semantic data from their regular database items.

An entirely different and highly visible LOD application is the BBC Music service Beta <http://www.bbc.co.uk/music/artists>. The backend of this BBC project comes from the Linked Data world - specifically MusicBrainz, an open content music "metadatabase" that lists information for over 400,000 artists.

<http://blog.marchibbins.com/2009/02/18/i-found-out/>

<http://blogs.talis.com/nodalities/2009/01/building-coherence-at-bbccouk.php>

Furthermore, BBC is republishing its own aggregated RDF as a service to the community.



Relevance to LarKC: Take-up of Semantic Web technology in various commercial non-IT markets is an important sign of maturity for potential deployment opportunities for LarKC.

3.3. Data.gov and Data.gov.uk

A special case of LOD are the datasets being brought online by the US and UK Governments as part of their drives towards more open and transparent government. The following are just some of the reports of activities in this sphere:

- Early description of the UK programme are at <http://www.silicon.com/management/public-sector/2009/06/10/brown-enlists-web-father-to-open-up-govt-data-39440349/> and <http://www.computing.co.uk/computing/news/2244720/tim-berners-lee-gives-first>
- In the short term (2010), the UK Government will publish 2,000 datasets: these will contain laws, traffic counts, property prices, motoring offences, and the top ranking criminal offences per county. <http://3.ly/xad>
- In the report at <http://i5.be/akg> the UK Gov promises "1000's of LOD sets" on transport, health service data, maps, crime statistics, educational data. The report contains a pledge to have "the majority of government data as Linked Open Data by June 2011.
- Following the national initiative, many local authorities in the UK are putting data on-line. Examples from London are <http://data.london.gov.uk>, and <http://www.guardian.co.uk/technology/2010/jan/06/london-datastore-launch-johnson-mashups>
- The UK Government has published guidelines on its data policies data: <http://www.hmg.gov.uk/frontlinefirst/action1/transparency.aspx>
- All UK central government consultation descriptions must use RDF/a, starting January 2010. <http://coi.gov.uk/guidance.php?page=315#section3d>
- Websites are beginning to emerge where the government data is beginning to be deployed. Some examples are <http://www.thisweknow.org/> and <http://data-gov.tw.rpi.edu>

Relevance for LarKC: Unexpected at the time of LarKC conception, it now seems that governmental data is rapidly becoming one of the best showcases of semantic web technology. None of the LarKC use-cases is currently concerned with this data.

3.4. Linked Open Drug Data

Linking Open Drug Data (LODD) is a further special case of LOD, aiming specifically at linking the various sources of drug data together to answer interesting scientific and business questions <http://esw.w3.org/topic/HCLSIG/LODD> . The latest public figures are from August 2008, which say that the data sets consist of over 8 million RDF triples, which are interlinked by more than 370,000 RDF links (As of August 2009). A highlight of this project is using state-of-the-art semantic link discovery techniques for interlinking the published datasets (since interlinking datasets is a weak part of the general LOD, with on average only 1 link between datasets for every 1000 triples in the datasets).

The closely related Bio2RDF project has published 40 biology-, gene- and medical-related datasets (altogether 2.3 billion triples).

Relevance for LarKC: These datasets are directly relevant to the use-cases in WP7a and WP7b, and in fact LarKC staff from Astrazeneca is involved in the LODD effort.



3.5. Deployment of GoodRelations

Although not part of LOD, the GoodRelations ontology has seen take-up by substantial commercial players: BestBuy.com, the largest specialty retailer of consumer electronics in the United States and accounting for 19% of the market, has published its entire catalogue in RDF. With 450,000 individual items and about 60 triples per item, this totals to about 27 million RDF triples.

<http://tinyurl.com/bestbuy-goodrelations>

<http://ebusiness-unibw.org/pipermail/goodrelations/2009-August/000103.html>

Furthermore, there are early signs that the inclusion of this RDF as RDF/a on the GoodRelations product pages seems to have a large impact on Google ranking, with a reported 30% traffic increase and 15% increase of clickthrough rates

<http://ebusiness-unibw.org/pipermail/goodrelations/2009-December/000152.html> .

Similarly, the book publisher O'Reilly is now using the GoodRelations ontology to mark-up its product page <http://bit.ly/621yM3>

Relevance to LarKC: As before, take-up of Semantic Web technology in various commercial non-IT markets is an important sign of maturity for potential deployment opportunities for LarKC.

3.6. Wolfram Alpha

Although not strictly (or even: not at all) semantic web technology, the launch of Wolfram Alpha received a lot of interest. The service aims to provide answers to general knowledge questions hosting a large variety of datasets. Although these datasets have often been obtained from public sources, they are nevertheless locally hosted, carefully curated, and not publically accessible. For the moment, the strengths of Wolfram Alpha seem to be mostly in answering quantitative, technical and numerical questions. It is as yet decidedly weak in terminological matters (e.g. not recognising that "The Queen of The Netherlands" is the same concept as "the head of state of The Netherlands").

Relevance for LarKC: Wolfram Alpha could become an interesting player in the Semantic Web landscape if it would provide a standards-compliant interface to its services (e.g. a SPARQL interface). In such a scenario, it could be easily absorbed as a plugin in LarKC pipelines. There are currently no signs of such a development.

3.7. New York Times tags

The New York Times announced that it will publish its collection of subject headings (in use for over 150 years to index NYT articles) as LOD data.nytimes.com. Ultimately, 30K tags in daily use at NYT will be in LOD format, linked to DBpedia & Freebase, CC-licensed. The first version, announced in October contained 5K tags. <http://blog.programmableweb.com/2009/12/08/new-york-times-releases-5000-tags-to-the-linked-data-cloud/> An update in January <http://bit.ly/7q9HGm> announces the expansion with a further 5K tags, totalling 10K tags.

Relevance for LarKC: A likely major source of meta-data terms (not data, but descriptive terms). Since there are published in LOD format, and LarKC is well placed to absorb LOD sources, the NYT vocabulary can be easily used in LarKC deployments.

3.8. Library of Congress name-authority file

After a very turbulent process, the Library of Congress has finally decided to publishing its name-authority file as LOD. This file contains authority records for thousands of government agencies in



the U.S., the U.K., and many other nations (as well as records for thousands of individuals), and hence would be an authoritative source of LOD instances. <http://bit.ly/6kHw8s>

Relevance for LarKC: As before, any growth in quality and quantity of LOD is good for LarKC.

3.9. Freebase

Freebase, which can be seen as a special case of community-constructed LOD announced the publication of its 10 millionth topic, up from 4 million a year ago, amounting to a doubling time of just under a year. <http://blog.freebase.com/2009/11/24/10-million-topics/>

Relevance for LarKC: since the publication of the Freebase data as an RDF dump, this is an increasingly important dataset for LarKC deployers.

3.10. Microsoft patent

A constant source of discussion (and even concern) in the Semantic Web community is the absence of much visible activity by Microsoft on semantic technologies. It was therefore potentially significant Microsoft put out a patent claim for SPARQL to SQL conversion <http://3.ly/QoQw> . The interest of this is not so much in the value of the patent claim (the community has already pointed out much prior art), but is rather of interest as a sign of activity within Microsoft.

Relevance for LarKC: Microsoft should remain on the radar for market activity observation, despite the lack of external visible activity.

3.11. LindedGeoData.org

Linkedgeodata.org announced their official launch. They made available both a large dataset derived from the OpenStreetMap database, converted this into RDF (which will contain around 2B statements, when the data conversion will be completed) and a set of REST/linked data services for online access to the dataset.

Relevance for LarKC: This is another sign of the growing interest for the location-based services on the Web and the data integration based on geo-spatial information.

3.12. Companies hosting Semantic Web data

Talis (a Semantic Web infrastructure and services provider) announced their new Connected Commons scheme, where they offer to host any RDF dataset (up to 50M triples) for free, and allow public SPARQL access to it, as long as that dataset is licensed under a creative commons or open data license <http://blogs.talis.com/n2/cc>.

Similarly, OpenLink is publically hosting of the lion share of all LOD datasets, both as a SPARQL endpoint and with a full-text index. <http://www.openlinksw.com/dataspace/kidehen@openlinksw.com/weblog/kidehen@openlinksw.com/s%20BLOG%20%5B127%5D/1539>

Amazon is hosting a large number of public datasets, much of which is also part of LOD <http://aws.amazon.com/publicdatasets/>. These are accessible under Amazon's commercial EC2 compute Service.



Relevance for LarKC: these hosting services (both commercial and public) greatly increase the opportunities for LarKC deployment. The current hosting practices by Talis, OpenLink and Amazon fully comply with standards that also underlie LarKC.

3.13. Business models for Semantic technologies

The community is beginning to think explicitly about business models. The VALUE-IT project <http://www.value-it.eu/> is a rich source of information, including their workshop "Market Analysis for Semantic Technologies" (listed at <http://www.estc2009.com/program/workshops>). Worth mentioning is the VALUE-IT report on Semantic Technologies for the Enterprise (STE), at <http://i5.be/am2>) and a summary of its key-findings at <http://blog.value-it.isoco.net/?p=243> .

The recent book "Pull: The Power of the Semantic Web to Change Business" by D. Siegal explicitly addresses the business impact of semantic technologies. See <http://blogcritics.org/books/article/book-review-pull-the-power-of/> for a review.

LOD is currently suffering from a serious lack of reward models. (currently the only reward for publishers of successful LOD sets is that their servers get overloaded). One of the first attempts to start thinking about business models for Linked Data can be found at <http://www.chiefmartec.com/2010/01/7-business-models-for-linked-data.html>

There has been some excitement about the possibilities of semantic technologies for marketing purposes <http://www.chiefmartec.com/2008/03/marketing-in-th.html>

Relevance for LarKC: This emerging line of line of thinking in the community should directly feed into exploitation plans for LarKC developed in WP9.

3.14. Search engines

Although semantic web technologies have not yet made a major impact on the search technology market, there have been some developments. The most visible of this has been the announcement that Google is deploying harvest RDF/a in its rich snippets service, following the trail open by Yahoo a year earlier. (<http://googlewebmastercentral.blogspot.com/2009/05/introducing-rich-snippets.html>)

There are also a number of smaller start-ups that aim to deploy semantics in search. A recent and intriguing entry is <http://kngine.com/>; we will not aim to give an exhaustive survey here.

3.15. Drupal supporting RDF/a

Drupal is one of the major content management platforms on the Web, with hundreds of thousands of installations. Drupal's most recent release includes support for publishing RDF/a that is automatically generated out of the slot/value representation that is used in the CMS engine. With sites as whitehouse.gov run by Drupal this could provide a substantial push towards the data volumes published in RDF/a, which in turns feeds into Google rich snippets, and thus providing an important stimulus to the entire semantic web ecosystem. <http://buytaert.net/rdfa-and-drupal>
<http://drupaleasy.com/blogs/ultimike/2009/06/rdf-drupal-what-can-i-do-today-drupal-6-semantic-web>

There have been recent announcements MediaWiki (the most frequently used Wiki platform) also supporting RDF/a (or at least of reviving the efforts to do so).



Relevance for LarKC: these developments make it urgent for LarKC to have good RDF/a TRANSFORMer plugins.

3.16. Triple stores

The market for triple stores is continuing to evolve rapidly, with new stores (or new versions of stores) coming on the market almost monthly. Most stores are reasonably standards compliant, and there is much emphasis on performance and scalability. Unfortunately, comparisons are hampered by unclear performance characteristics (e.g. 5store recently mentioning 1 billion triples, but unclear at which cluster size, what queries are supported, how query response time degrades with query complexity, if any forms of inference are supported, etc). Benchmarking by Ontotext indicates that the LarKC data layer is still very well placed wrt its competitors.

Relevance for LarKC: The LarKC data layer is still well placed in the market. Standard compliance by LarKC and store vendors make it in principle possible to implement LarKC workflows on top of stores by other vendors.

4. Conclusion

4.1. Conclusion on Standards

In the area of standards, the steady developments are very well aligned with the technical requirements of LarKC. The most important one of these is currently the SPARQ effort. The work in the SPARQL WG on describing SPARQL endpoints, on updates to RDF graphs, on a SPARQL protocol and on querying under entailment regimes are all needed by LarKC and the LarKC architecture is already precluding these new developments. Second in importance is the RDF/A development, not so much because of its technical role in LarKC, but rather because it promises to be an important device bringing more RDF triples on-line. Less crucial, but still well aligned are OWL2 and RIF, with RIF's non-acceptance of RDF as a somewhat annoying detail.

The overall conclusion here is that standards development is progressing steadily, and remains well aligned with LarKC's needs and goals.

4.2. Conclusions on Market Observations

In contrast, the market developments have been much more turbulent and unpredictable. In the past 2 years (and mostly in the past year), Linked Open Data has established itself firmly as the main breakthrough possibility for Semantic Web technology. This was far from obvious as little as 2 years ago, when a host of different scenario's were being pursued: semantic search, semantic enterprise data, semantic annotation of existing web-pages, semantic web-services, and others. It would now seem that none of these are going to provide the breakthrough for semantic technologies, but that LOD is the best bet for this.

A second unpredicted development has been the rise of government data as one of the major and publically most visible applications of Linked Open Data. Besides the life-science data-sets which are mostly aimed at a scientific audience, government data is rapidly becoming the 2nd largest source of LOD, and is certainly attracting most public attention.

A major drawback of LOD is its current lack of any reasonable business scenario (whereas all the previously mentioned directions did come with credible business models). It is currently hard to predict whether this lack of business models for LOD will become a major obstacle or if such models will be developed in the short term.



The rise of LOD is certainly a positive development for LarKC (although as said, mostly unforeseen at the time of LarKC's inception). The LOD models produces very large amount of publically available semantic web data (much more so than many of the previously fashionable scenario's). This growth produces a strong need for very large scale distributed processing, ranging from basic querying to advanced reasoning. Furthermore, by its very nature, LOD data is highly heterogeneous, incomplete, and often internally inconsistent. This fits very well with LarKC's emphasis on approximation and anytime behaviour as important ways to achieve scalability.

In short: the rise of LOD as a dominant model for the Semantic Web is well aligned with LarKC's goals and technology; the lack of business models is a threat to the success of LOD (and LarKC) in the medium term.