LarKC

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

FP7 – 215535

D8.3

Community-building Efforts and Cross-fertilization

Coordinator: Zhisheng Huang (VUA)
With contributions from: Frank van Harmelen (VUA), Christoph Fuchs (STI), Michael Witbrock (CycEur), Alexey Cheptsov(HLRS), Tony Lee(Saltlux)
Quality Assessor: Hansjörg Neth (MPG)
Quality Controller: Christoph Fuchs (STI Innsbruck)

Document Identifier:	LarKC/2008/D8.3
Class Deliverable:	LarKC EU-IST-2008-215535
Version:	version 4.0.0
Date:	September 20, 2011
State:	Final
Distribution:	Public

Executive Summary

This document reports the LarKC community building and cross-fertilization activities undertaken from month 34 to month 42 of the LarKC project. In particular, this version elaborates on the LarKC community building channels, on the cooperation with other projects and communities of interest, and presents up-to-date statistics of related LarKC work. In this document, we describe the cooperation with other projects, report organized events and upcoming events as community building efforts and cross-fertilization, and present the plan for exploiting LarKC after the LarKC project is finished.

Document Information

IST Project Number	FP7 – 215535	**Acronym**	LarKC
Full Title	Large Knowledge Collider		
Project URL	`http://www.larkc.eu/`		
Document URL			
EU Project Officer	Stefano Bertolo		

Deliverable	**Number**	8.3	**Title**	Community-building Efforts and Cross-fertilization
Work Package	**Number**	8	**Title**	Training, dissemination, community building, cross-fertilization

Date of Delivery	**Contractual**	M42	**Actual**	30-September-2010
Status	version 4.0.0		final ☒	
Nature	prototype ☐ report ☒ dissemination ☐			
Dissemination Level	public ☒ consortium ☐			

Authors (Partner)	Zhisheng Huang(VUA), Frank van Harmelen (VUA), Christoph Fuchs (STI), Michael Witbrock (CycEur),Alexey Cheptsov(HLRS), Tony Lee(Saltlux)			
Resp. Author	Zhisheng Huang (VUA)		**E-mail**	huang@cs.vu.nl
	Partner	VUA, STI, Saltlux, CycEuro	**Phone**	+31 (20) 5987823

Abstract (for dissemination)	This document reports the LarKC community building and cross-fertilization activities undertaken from month 34 to month 42 of the LarKC project. In particular, this version elaborates on the LarKC community building channels, on the cooperation with other projects and communities of interest, and presents up-to-date statistics of related LarKC work.
Keywords	community-building, cross-fertilization

Version Log			
Issue Date	**Rev No.**	**Author**	**Change**
May 17, 2011	0.1	Zhisheng Huang	Updated from Month 33 deliverable
July 25, 2011	0.2	Zhisheng Huang	Updated Chap 3, Chap 4 and Chap 5
July 25, 2011	0.3	Frank van Harmelen	Updated Chap 3, Chap 4 and Chap 5
August 11, 2011	0.4	Michael Witbrock	Update
August 11, 2011	0.5	Zhisheng Huang	Update
Sept 7, 2011	0.6	Zhisheng Huang	Revision
Sept 14, 2011	0.7	Hansjörg Neth	Minor revisions in the context of quality control
Sept 19, 2011	0.8	Zhisheng Huang	Final Revision

Project Consortium Information

Acronym	Partner	Contact
Semantic Technology Institute Innsbruck http://www.sti-innsbruck.at		Prof. Dr. Dieter Fensel Semantic Technology Institute (STI) Innsbruck, Austria Email: dieter.fensel@sti-innsbruck.at
AstraZeneca AB http://www.astrazeneca.com		Bosse Andersson AstraZeneca Lund, Sweden Email: bo.h.andersson@astrazeneca.com
CEFRIEL SCRL. http://www.cefriel.it		Prof. Dr. Emanuele Della Valle CEFRIEL SCRL. Milano, Italy Email: emanuele.dellavalle@cefriel.it
CYCORP, RAZISKOVANJE IN EKSPERIMENTALNI RAZVOJ D.O.O. http://cyceurope.com		Dr. Michael Witbrock CYCORP, RAZISKOVANJE IN EKSPERIMENTALNI RAZVOJ D.O.O., Ljubljana, Slovenia Email: witbrock@cyc.com
Höchstleistungsrechenzentrum, Universitaet Stuttgart http://www.hlrs.de		Matthias Assel Höchstleistungsrechenzentrum, Universität Stuttgart Stuttgart, Germany Email : assel@hlrs.de
MAX-PLANCK GESELLSCHAFT ZUR FOERDERUNG DER WISSENSCHAFTEN E.V. http://www.mpib-berlin.mpg.de		Dr. Lael Schooler, Max-Planck-Institut für Bildungsforschung Berlin, Germany Email: schooler@mpib-berlin.mpg.de
Ontotext Lab, Sirma Group Corp. http://www.ontotext.com		Atanas Kiryakov, Ontotext Lab, Sirma Group Corp. Sofia, Bulgaria Email: atanas.kiryakov@sirma.bg
SALTLUX INC. http://www.saltlux.com/EN/main.asp		Kono Kim SALTLUX INC Seoul, Korea E-mail: kono@saltlux.com
SIEMENS AKTIENGESELLSCHAFT http://www.siemens.de		Dr. Volker Tresp SIEMENS AKTIENGESELLSCHAFT München, Germany Email: volker.tresp@siemens.com
THE UNIVERSITY OF SHEFFIELD http://www.shef.ac.uk		Prof. Dr. Hamish Cunningham THE UNIVERSITY OF SHEFFIELD Sheffield, UK Email: h.cunningham@dcs.shef.ac.uk
VRIJE UNIVERSITEIT AMSTERDAM http://www.vu.nl		Prof. Dr. Frank van Harmelen VRIJE UNIVERSITEIT AMSTERDAM Amsterdam, Netherlands Email: Frank.van.Harmelen@cs.vu.nl
THE INTERNATIONAL WIC INSTITUTE, BEIJING UNIVERSITY OF TECHNOLOGY http://www.iwici.org		Prof. Dr. Ning Zhong THE INTERNATIONAL WIC INSTITUTE Mabeshi, Japan E-mail: zhong@maebashi-it.ac.jp
INTERNATIONAL AGENCY FOR RESEARCH ON CANCER http://www.iarc.fr		Dr. Paul Brennan INTERNATIONAL AGENCY FOR RESEARCH ON CANCER Lyon, France Email: brennan@iarc.fr
INFORMATION RETRIEVAL FACILITY http://www.ir-facility.org		Dr. John Tait INFORMATION RETRIEVAL FACILITY Vienna, Austria Email: john.tait@ir-facility.org

TECHNICAL UNIVERSITY OF CLUJ-NAPOCA http://www.utcluj.ro		Prof. Dr. Eng. Sergiu Nedevschi TECHNICAL UNIVERSITY OF CLUJNAPOCA Cluj-Napoca, Romania Email: sergiu.nedevschi@cs.utcluj.ro
SOFTGRESS S.R.L. http://www.softgress.com		Dr. Ioan Toma SOFTGRESS S.R.L. Cluj-Napoca, Romania Email: ioan.toma@softgress.com

Table of Contents

List of Figures

1 Introduction

The primary objective of LarKC WP8—entitled *Training, dissemination, community building and cross-fertilisation*—is to ensure that the work carried out throughout the project is aligned with similar inter-disciplinary initiatives in order to increase its quality, impact and visibility within the relevant communities of practice. A core component of the activities described in this document contain community building and networking efforts. These include activities targeted at encouraging inter-disciplinary scientific collaboration and cross-fertilization, as well as activities aimed towards promoting the creation of a user community of early adopters of and contributors to LarKC technologies. The former collaboration and cross-fertilization activities are achieved by individual partners through their collaboration with relevant initiatives and parties in their field of activities, and through the organization of dedicated events. The main instrument for the implementation of a community of users and early adopters is the Early Access Group.

This document is a version of an accumulated report of the LarKC community building and cross-fertilization activities throughout the duration of the project. The corresponding materials will be regularly updated based on the user feedback received. The results of this interaction are summarized in the versions of D8.3, which were or are due in M6, M18, M33 (reported in previous versions of this document), and M42 (in the current version) of the LarKC project, respectively.

The current version of this document reports on relevant activities undertaken from Month 34 until Month 42 (January 2011 to September 2011) of LarKC. In particular, it elaborates on the LarKC community building channels, on the cooperation with other projects and communities of interest, and presents up-to-date statistics of LarKC work. In this document, we describe the cooperation with other projects, report organized events and upcoming events as community building efforts and cross-fertilization, and present the plan for exploiting LarKC after the LarKC project is finished.

2 LarKC Communication Channels

In this chapter, we will provide an overview of the various community building, networking, and cross-fertilization channels implemented in the LarKC project. These channels comprise the LarKC web site, the project mailing lists, the LarKC Wiki, and the LarKC Weblog. These channels also support complementary activities, such as training and dissemination, which are addressed in deliverables D8.1 and D8.2, respectively.

2.1 LarKC Web Site

The project web site, available at `http://www.larkc.eu`, functions as a project dissemination tool, as a training portal, and as a place demonstrating the most important achievements of the project. It is based on a content management system, so that project participants and interested users can register and be notified when new deliverables or software releases become available.

A screenshot of the LarKC web site is shown in Figure 2.1. Additional information about the project web site is available in deliverable D8.2.

Figure 2.1: The LarKC web site.

2.2 LarKC Mailing Lists

The LarKC mailing lists are designed for the communication by emails for internal groups or external groups (e.g., early access group). At this time, we have set up the following mailing lists, mainly for the internal group communications:

- **Ict-larkc** for discussions of the whole LarKC consortium
- **Larkc-epmb** for discussions of the LarKC Executive Project management Board
- **Larkc-tmb** for discussions of the LarKC Technical Management Board
- **Larkc-wp1** for internal discussions of LarKC WP1
- **Larkc-wp2** for internal discussions of LarKC WP2
- **Larkc-wp3** for internal discussions of LarKC WP3
- **Larkc-wp4** for internal discussions of LarKC WP4
- **Larkc-wp5** for internal discussions of LarKC WP5
- **Larkc-wp6** for internal discussions of LarKC WP6
- **Larkc-wp7a** for internal discussions of LarKC WP7a
- **Larkc-wp7b** for internal discussions of LarKC WP7b
- **Larkc-wp8** for internal discussions of LarKC WP8
- **Larkc-wp9** for internal discussions of LarKC WP9
- **Larkc-wp11** for internal discussions of LarKC WP11
- **Larkc-whitepaper** for internal discussions of LarKC Whitepaper(s)

Furthermore, we have created the following mailing lists for communications with partners external to the LarKC consortium:

- **larkc-early-adopters** a mailing-list for discussions of the early adopters group
- **larkc-dev@googlegroups.com** a Google group for discussions of the LarKC developers
- **larkc-user-support@lists.sourceforge.net** a SourceForge group for the LarKC users

2.3 LarKC Wiki

The LarKC Wiki, available at `http://wiki.larkc.eu`, serves as a project discussion forum and knowledge sharing platform, and also as a public portal about relevant information. A screenshot of the LarKC Wiki is shown in Figure 2.2.

The LarKC wiki consists of several parts (see `http://wiki.larkc.eu/LarkcProject` and below). In the workpackage part, each work package created their own sub-section with pages on relevant topics, work scheduling, information sharing, etc. The management part includes project-internal management information, local organization details of project meetings, etc. Those parts are accessible to project members (both for read- and write-access). All members can see and edit all information. They are indeed encouraged to do so.

The LarKC wiki is readable for everyone, but writable only for project partners. It contains information about various relevant technology topics (such as approximate reasoning, triple stores, and distributed computing), giving a general introduction to these topics and links to further information. It also lists information about LarKC events, early adopters, and opportunities to meet and interact with LarKC project members.

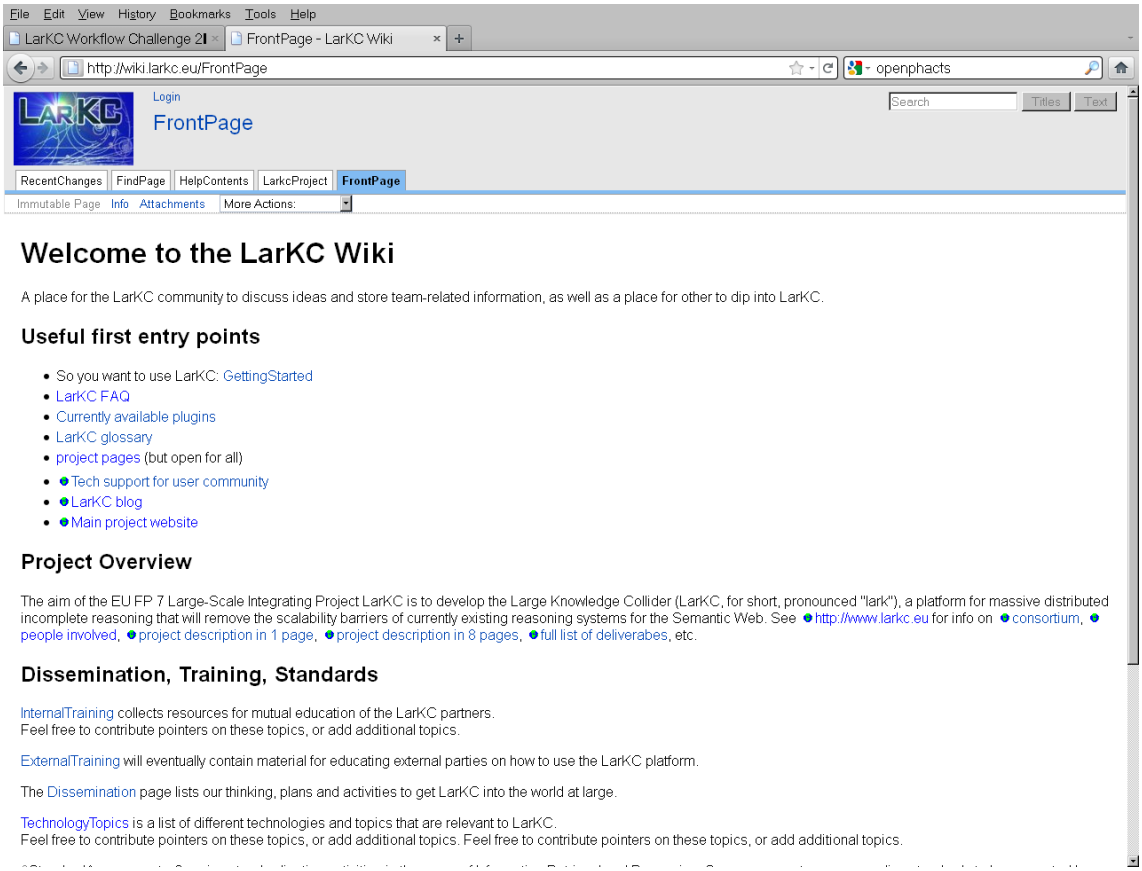

Figure 2.2: The LarKC Wiki.

2.4 LarKC Blog

The LarKC Blog, available at `http://blog.larkc.eu`, serves as a place for knowledge sharing that is accessible both for internal members and external visitors. A screenshot of the LarKC Blog is shown in Figure 2.3.

Figure 2.3: The LarKC Blog.

At the time of writing (July 25, 2011), the LarKC Blog contains 152 posts and 122 comments, grouped into 19 categories and described by 148 tags, which cover a diverse range of topics. Figure 2.3 shows some of the topics that are posted on the LarKC Blog. So far, these pages have been viewed 367,252 times by 64,845 visitors and 136,864 feeds since June 5th 2008, when the first message was posted to the Blog. The overview statistics of the LarKC Blog are shown in Figure 2.4.

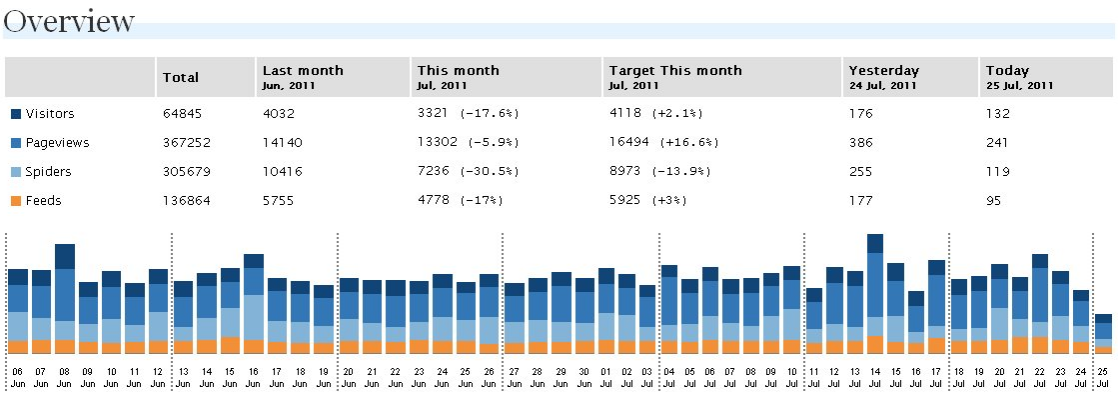

Figure 2.4: Overview statistics of the LarKC Blog.

2.5 LarKC User Forum

2.5.1 LarKC Developer Forum

As described in Deliverable 8.1, delivered in month 33, the LarKC developer forum provides a location to early adopters and developers for the discussion of issues related to the LarKC platform. It provides access to forums, trackers, and mailing lists in which those developing the LarKC platform, those implementing LarKC plugins, and those creating LarKC workflows can discuss issues related to using the new LarKC technology. The developer forum `http://sourceforge.net/projects/larkc/develop` is currently hosted within the SourceForge environment, an open source platform `sourceforge.net` (see Figure 2.6), This ensures greater visibility of the LarKC developer forum. A screenshot of the current LarKC developer forum at the SourceForge is shown Figure 2.5.

2.5.2 LarKC Chinese User Forum

The LarKC Chinese User Forum is based on the forum of W3C-China, available at `http://semanticweb.org.cn/list.asp?boardid=80`, serves as an external project discussion forum and knowledge sharing platform for Chinese users/developers. A screenshot of the LarKC Chinese Forum is shown in Figure 2.7.

The LarKC Chinese User Forum sends news update related to LarKC on a monthly basis, plus optional information in case of some major event, such as an early adopters workshop. This forum provides release-related information to the Chinese user community. Based on the translation functionality provided by Google groups, the forum is also responsible for translating development issues written in Chinese to the LarKC Developers Forum which supports English inquiries. Major issues proposed by LarKC Chinese users have been collected and sent to LarKC consortium or discussed on the LarKC Chinese Developer Workshop. The Chinese version of major LarKC documents will be released through this forum (e.g., the Getting Started documents, etc.).

At the time of writing (August 2011), the LarKC Chinese User Forum contains 91 posts to discuss 37 topics, which range from the runtime problems of the LarKC platform to LarKC community activities, such as the LarKC PhD Symposium.

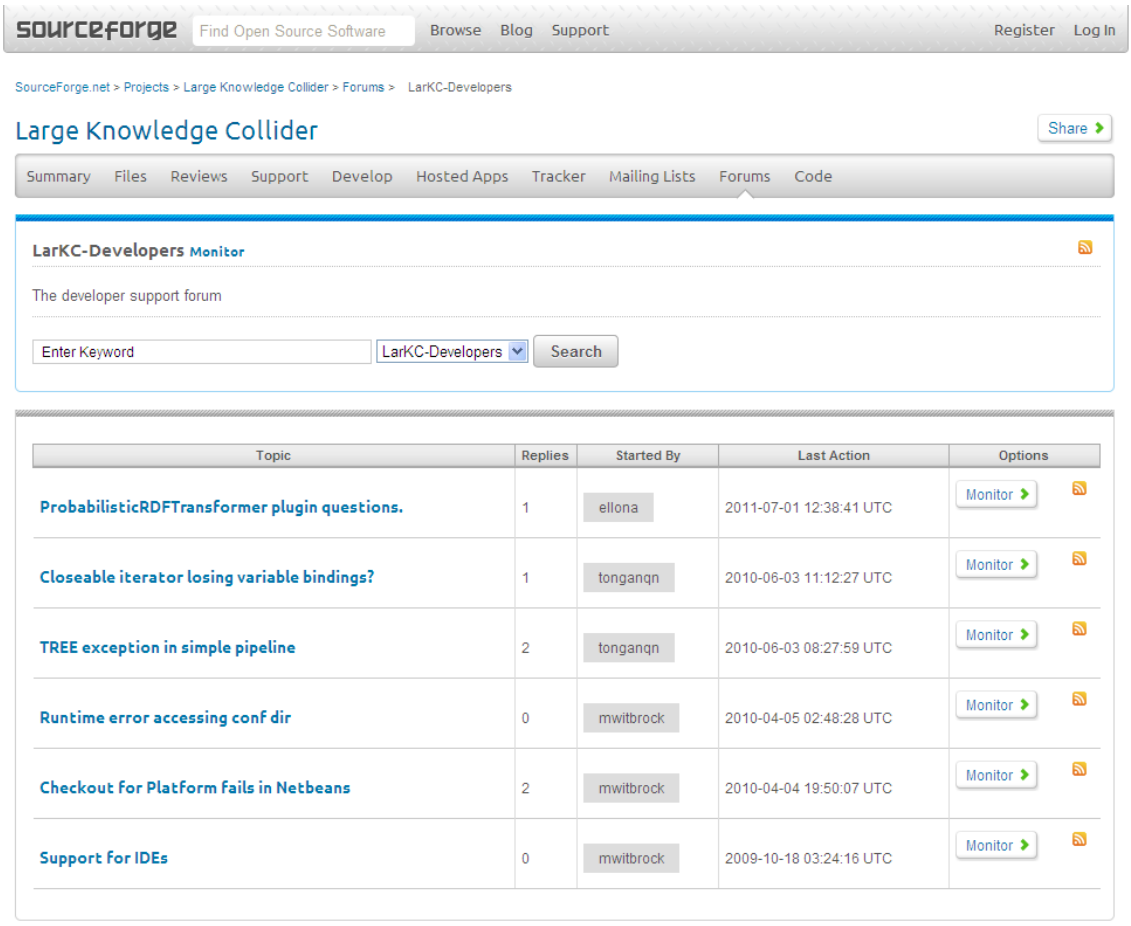

Figure 2.5: The LarKC Developer Forum at SourceForge.

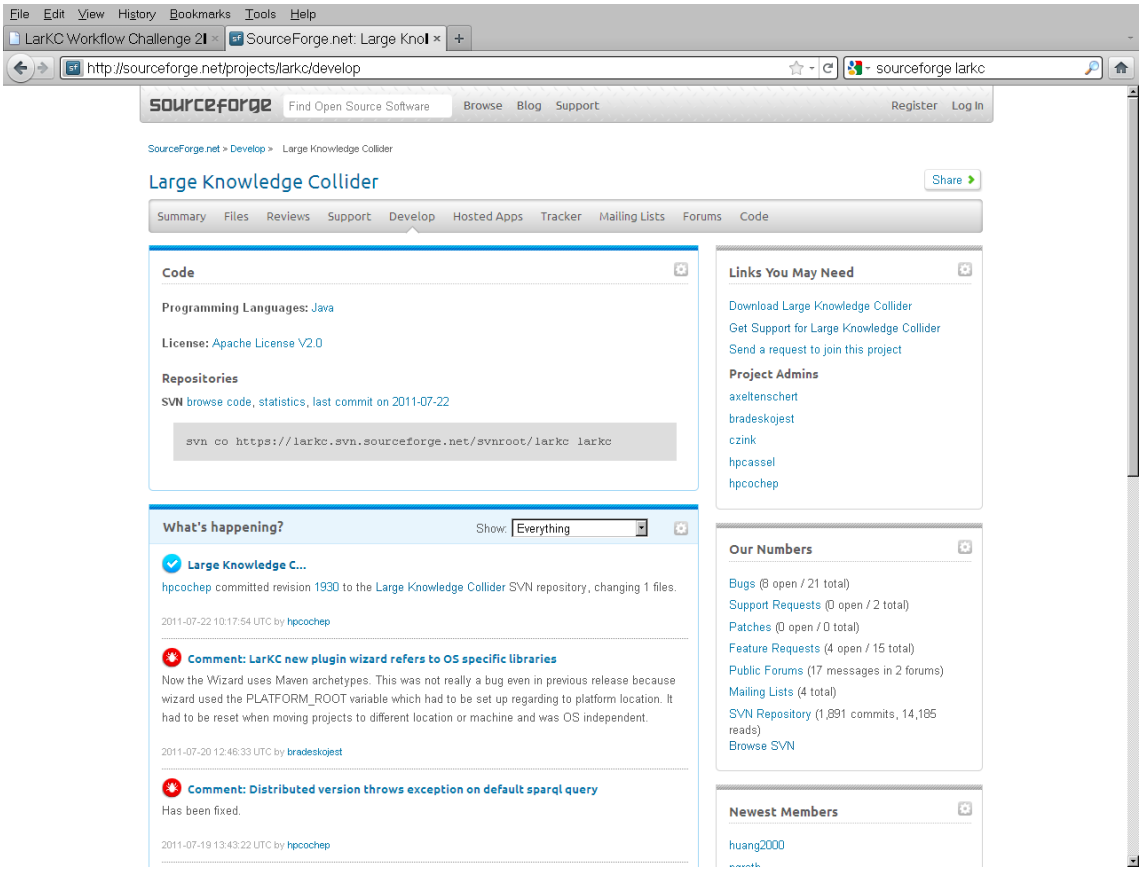

Figure 2.6: The LarKC platform at SourceForge.

2.6 LarKC Tweets

At the time of writing (July 2011) there are over 60 tweets from LarKC. Their audience exceeds 1.000 followers, including tweets and re-tweets. The LarKC tweets appear in the top-100 Semantic Web twitter members.

Figure 2.7: The LarKC Chinese user forum.

3 Cooperation and Advisory Board

3.1 Cooperation with Other Projects

LarKC has initiated a cross-project cluster on Web scalability within STI and several related projects of the current and future EU FP already expressed an interest to the contribution. These projects share the overall goal of developing new approaches for realizing scalable systems at Web level using semantic technologies and will benefit from the support provided by STI through its road mapping, standardization, commercialization, and education services.

SEALS (`http://www.seals-project.eu/`) is a project aiming for Semantic Evaluation at Large Scale. The goal of the SEALS project is to provide an independent, open, scalable, extensible and sustainable infrastructure. The SEALS platform allows the remote evaluation of semantic technologies, thereby providing an objective comparison of the different existing semantic technologies. This will allow researchers and users to effectively compare the available technologies, helping them to select appropriate technologies and advancing the state of the art through continuous evaluation. The SEALS platform will provide an integrated set of semantic technology evaluation services and test suites. They will be used in two public and worldwide evaluation campaigns. The results of these evaluation campaigns will be used to create semantic technology roadmaps identifying sets of efficient and compatible tools for developing large-scale semantic applications. LarKC cooperates with SEALS to integrate the evaluation methods and benchmarks developed in the context of LarKC with the SEALS platform. SEALS and LarKC co-organized the International Workshop on Evaluation of Semantic Technologies (IWEST 2010, `http://oeg-lia3.dia.fi.upm.es/iwest2010/`), which was co-located with the 2010 International Semantic Web Conference, held in Shanghai, Nov. 2010.

OpenPhacts (`http://www.openphacts.org/`) Open PHACTS (Open Pharmacological Concepts Triple Store) is a knowledge management project of the Innovative Medicines Initiative (IMI), a unique partnership between the European Community and the European Federation of Pharmaceutical Industries and Associations (EFPIA). The Open PHACTS consortium will create an open innovative platform, Open Pharmacological Space, which will be freely accessible for knowledge discovery and verification. It will also serve other IMI projects, the broader pharmaceutical industry, and other public drug discovery efforts. Open PHACTS will provide a growing body of data on small molecules, their pharmacological profiles, pharmacokinetics, ADMET data, biological targets and pathways in a semantically interoperable format. Aligning and integrating proprietary and public data sources into a single system is currently a very difficult and time consuming task, repeated across companies, institutes and academic laboratories. Open PHACTS will develop an integrated system, guided by typical research questions from participating drug discovery teams. The Open PHACTS consortium comprises 14 European core academic and SME partners and 8 EFPIA members. VUA is one of the partner in the Consortium. The OpenPhacts consortium has adopted LarKC as the core infrastructure for their first prototype. Developers from a dozen OpenPhacts partners have been developing LarKC plugins. All of

the major components of the first OpenPhacts prototype have been connected through LarKC, enabling to solve one of the OpenPhacts challenge problems ("find oxodoreductase inhibitors with an IC50 values below a threshold along with species info"). This is done through integrating Brenda, Enzyme, PDSP, Metadata from Uniprot, Kegg all as LarKC plugins. Ongoing work concerns chemical similarity search from a sparql query using chemspider web services, the use of PathVisio, Utopia and Chemical Structure Info user interfaces on top of the LarKC platform, and an Identity Resolution Service leveraging BridgeDB and Concept Wiki.

LOD2 (`http://lod2.eu/`) LOD2 is an FP7-funded IP, aiming to build enterprise-ready tools and methodologies for exposing and managing very large amounts of structured information on the Data Web; a testbed and bootstrap network of high-quality multi-domain, multi-lingual ontologies from sources such as Wikipedia and OpenStreetMap; algorithms based on machine learning for automatically interlinking and fusing data from the Web; standards and methods for reliably tracking provenance, ensuring privacy and data security as well as for assessing the quality of information; and adaptive tools for searching, browsing, and authoring of Linked Data. In order to achieve these goals, LOD2 has been looking at LarKC as a potential engineering platform. Discussions have taken place between technical members of both consortia. One of the issues currently under technical investigation is the possibility to make LarKC independent from the current implementation of the data layer, in order to allow the use of other triple-store implementations. Work on this is in progress.

3.2 Advisory Board

The purpose of the Advisory Board is to provide the project with advice on all matters, e.g., technical issues, dissemination opportunities, possible links with industry and with other research projects, etc. Current members of the advisory board are:

- Dr. Mark Greaves, director, Knowledge Systems, Vulcan Inc.;
- Dr. Ron Brachman, worldwide head of research at Yahoo;
- Prof. Mark Musen, Head of the Center for Biomedical Informatics Research at Stanford University.

The LarKC Advisory Board meeting will take place on Sunday, 25-Sep-11, in London, UK. It will be a full-day meeting, starting at 09:30 and ending around 17h.

3.3 Cooperation with Researchers in Asia

South Korea: In 2009, Frank van Harmelen visited South Korea for two consecutive years. The visit was fully funded by Saltlux and the Korean government. He delivered a keynote on LarKC at a national symposium and met for a workshop with researchers at KAIST, the top South Korean research institute, who already have constructed massive ontologies along with other necessary tools as an infrastructure for industrial applications of semantic technologies. Korean

researchers and scientists expressed huge interest into LarKC for its scalability and flexibility. Also, van Harmelen introduced EU efforts and experiences in building legal and legislation ontologies, such like SEKT, BEST, MetaLex, ESTRELLA/LKif, as a response to a Korean government request.

LarKC researchers cooperated with the researchers of Seoul National University and organized the first Asian Workshop on Scalable Semantic Data Processing (AS2DP2009, `http://wasp.cs.vu.nl/workshop/as2dp2009`). The workshop was co-located with the 4th Asian Semantic Web Conference (`http://www.aswc2009.org`), in Shanghai, China.

The LarKC Consortium organized the Semantic Web Open Seminar (held on July 26, 2010) in Seoul. This was a half day event which consisted of 7 different sessions. More than 160 participants attended the semniar.

The seminar at Korea Advanced Institute of Science and Technology (KAIST) was held on July 27–28, 2010 in Daejon. There were 15 attendees participating in the discussions on future collaborations and active contributions to LarKC plugins. The seminar at Korea Institute of Construction Technology (KICT) was held on July 29, 2010 in Seoul, as a technical seminar with intensive discussions on possible collaborations. 13 people attended the KICT seminar.

China: The Chinese LarKC developer forum is a W3C-China forum group for discussions on various issues for Chinese LarKC developers and users. So far there are more than 20 Chinese researchers/developers who have shown their willingness to contribute to the development of LarKC. Of them, eleven researchers/developers (two professors, one lecturer, six PhD students, and some Masters students) belong to the WICI team in Beijing, five researchers/developers (one associated professor and four Masters students) are from the faculty of computer science, at the Beijing University of Technology, and five researchers/developers (one professor, one lecturer, and three Masters students) are from the Jiangsu University of Science and Technology.

Zhisheng Huang gave a series of lectures at Chinese universities in 2011, including several lectures at Jiangsu University of Science and Technology and Southeast University, Nanjing, eight lectures on the Semantic Web and LarKC in the 2011 Chinese Summer School in the Semantic Web, which attracted about 180 Chinese participants.

Malaysia: The first contact with Malaysian researchers/developers was established with the Knowledge Technology Cluster of MIMOS BERHAD in Kuala Lumpur. MIMOS is the premier research centre on frontier technologies in Malaysia. The MIMOS pioneers innovative information and communication technologies toward growing globally competitive indigenous industries. MIMOS is actively studying LarKC with the intention to become early adopters and to establish a long-lasting link. This collaboration shall be materialized and intensified in 2011.

3.4 LarKC in the World

Although LarKC is based in Europe, the project of building, and applying, web-scale reasoning is world wide. One of the most exciting things about living in a connected

world, and a world of abundant, location independent computational resources, is that people anywhere in the world can do world class AI research, and develop applications based on that research. The recent, and very rapid, increase in *internet bandwidth going into Africa*[1] means that one can now use Shazam[2] to get impromptu karaoke lyrics for the Texas country-and-western playing in a hotel bar in Accra. It also means that previously isolated African researchers can make a full contribution to the advance of semantic technology. In February, partially supported by the FP7 Active project[3], we had the opportunity to present LarKC, and the potential benefits of AI and human-computer collaboration, to students and researchers at the *Ghana-India Kofi Annan Centre of Excellence in ICT in Ghana*[4]. Discussion following the talks was lively, with great local ideas for the application of AI in knowledge capture from small farmers, and resource allocation for rural health care. Video from some talks is being made available on VideoLectures.net, there was good coverage from the local media, and we look forward to building a collaboration with our new colleagues.

3.5 Chinese LarKC Book

The China Higher Education Press will publish a LarKC book in Chinese. This book will appear in the book series of Web Intelligence and Web Science[5].

This Chinese LarKC book consists of two parts: Technology part and application part. The technology part covers the topics of LarKC platform, development guide and various plugins and workflows. The application part covers the topics of Linked Life Data, semantic information retrieval, urban computing, and cancer study. The main contributors of the book are six Chinese researchers in the LarKC Consortium, who are from Amsterdam, WICI, and Siemens. See the appended text below for the detail. The book is expected to be published by the end of this year.

Here is the outline of the book content and the main contributors:

```
Chapter 1 Introduction to LarKC
by Zhisheng Huang (VUA) and Ning Zhong (WICI)

Chapter 2 LarKC Platform
by Jun Fang (VUA)

Chapter 3 Identification  and Selection
by Yi Zeng (WICI)

Chapter 4 Abstraction and Transformation
by Yi Huang (SIEMENS)

Chapter 5 Reasoning  and Deciding
by Jun Fang (VUA) and Zhisheng Huang (VUA)
```

[1] http://www.moreintelligentlife.co.uk/content/ideas/jm-ledgard/digital-africa
[2] http://www.shazam.com/
[3] http://www.active-project.eu/
[4] http://en.wikipedia.org/wiki/Ghana-India_Kofi_Annan_Centre_of_Excellence_in_ICT
[5] http://www.wici-lab.org/wici/WIWS/

Chapter 6 LarKC Development Guide
by Zhisheng Huang (VUA) and Jun Fang (VUA)

Chapter 7 Linked Life Data
by Yi Huang (SIEMENS) and Zhisheng Huang (VUA)

Chapter 8 Semantic information retrieval for biomedical applications
by Ru He (SIEMENS) and Zhisheng Huang (VUA)

Chapter 9 Semantic Technology and Gene Study
by Zhisheng Huang (VUA)

Chapter 10 Urban Computing
by Yi Huang (SIEMENS) and Zhisheng Huang (VUA)

Chapter 11 Conclusions
by Zhisheng Huang (VUA), Ru He (SIEMENS), and Ning Zhong (WICI)

4 Events as Community Building Efforts and Cross-fertilization

4.1 Inter-disciplinary Community Building

The scientific communities that is targeted by our inter-disciplinary community building efforts are very diverse and hitherto largely independent of each other. The include cognitive systems, web intelligence, semantic web and knowledge technologies, machine learning, multi-agent systems, information and game theory. One way we promote cross-fertilization is through inter-disciplinary workshops, organized alongside major conferences in the respective research fields.

The LarKC partners representing each of these diverse research areas are responsible for promoting the project results within their respective communities:

- WICI for the Web Intelligence community, through the participation in the Web Intelligence Consortium;
- MPG and WICI for the cognitive science research community. Lael Schooler of MPG gave a talk on possible links between LarKC components and cognitive architectures at the 2009 ACT-R workshop (before the main International Cognitive Science conference, see `http://csjarchive.cogsci.rpi.edu/Proceedings/2009/tutorials.html`) and a similar presentation at the European Society for Cognitive Psychology's 2010 Summer School in Computational and Mathematical Modeling of Cognition (ESCOP, `http://perswww.kuleuven.be/~u0049679/ss2010/`), which took place on July 9–19, 2010, in Mallnitz, Austria;
- STI and VUA for the Semantic Web and knowledge technology community, through their extensive involvement as organizers of all major conferences etc. In this respect, LarKC partners are already involved in the organization of scientific workshops on topics which are integral part of the project.

4.2 Early Adopters Group

LarKC has created an Early Adopters Group that is open for participation to scientists working in related fields. Members of this group get early access to project infrastructure and are invited to experiment with plug-ins on the LarKC platform. This both verifies the generality of the plug-in interfaces and encourages the take-up of project results by teams external to the LarKC consortium.

The Scalability of Semantic Computing special session at the IEEE ICSC2008, held in Santa Clara, CA, USA, on August 4–7, 2008 (`http://icsc.eecs.uci.edu`), was a cross-project initiative of several FP6 and FP7 European projects which share the overall goal of developing new approaches for realizing scalable systems at Web level using Semantic Web and Semantic Web services technology. This cross-project coordination was lead by STI International and consisted of FP6 IP SUPER, FP6 STREP TripCom, FP7 IP LarKC, FP7 IP SOA4All, FP7 IP OKKAM, FP7 SA Service Web 3.0, and FP7 IP SEALS.

The special session, entitled "Scalability in Semantic Computing: the European View", provided an overview of some of the most important achievements of European research in its enterprise towards the realization of scalable and robust semantic technologies. The papers presented throughout the session addressed the following topics:

- Semantic Service Orientation – the FP7 IP SOA4ALL;
- Semantic Middleware – the FP6 STREP TripCom;
- Semantic Web Reasoning – the FP7 IP LarKC;
- Identity and Reference Management – the FP7 IP OKKAM;
- Semantic Business Process Management – the FP6 IP SUPER.

The event was rounded off by two further presentations: A keynote by Randy Shoup, Distinguished Architect in the eBay Marketplace Architecture Group, who presented real-world strategies for scaling, including partitioning, asynchrony, virtualization, and automation, applied at the largest auction site in the world, and potential use cases for semantic technologies and reasoning in this setting; and a concluding presentation by Graham Hench, STI International, who presented a first outline of a future conceptual road-map for scalable semantic computing leveraging the results, findings and insights of individual project-related contributions.

We organized the NeFors08 and NeFors10 (the 2008 and 2010 Workshop on New Forms of Reasoning for the Semantic Web: scalable, tolerant and dynamic (`http://nefors08.larkc.eu`). This workshops were intended to focus on the problems of scalability and robustness of reasoning on the Web, and furthermore to investigate alternative reasoning methods, which take incompleteness and distribution of data and knowledge as inherent properties into account. The workshop took place during the 3rd Asian Semantic Web Conference (ASWC2008), December 2008, in Thailand, as a full-day event. NeForS08 was the follow-up of the first international workshop on "New Forms of Reasoning for the Semantic Web: scalable, tolerant and dynamic", which took place in Busan, Korea, in November 2007 and was co-located with the 6th International Semantic Web Conference (ISWC 2007) and the 2nd Asian Semantic Web Conference (ASWC 2007). We organized the special issue on Web Scale Reasoning for the international journal of Semantic Computing (WebScaRE2009, `http://webscare.larkc.eu/`).

The following events have been organized to promote the community building efforts for LarKC:

- The second LarKC early adopters tutorial was co-located with ISWC 2009. This activity is considered as one of the most important events within 2009 for dissemination purposes.
- The first Asian Workshop on Scalable Semantic Data Processing (AS2DP 2009, `http://wasp.cs.vu.nl/workshop/as2dp2009`) was co-located with the 4th Asian Semantic Web Conference (`http://www.aswc2009.org`), in Shanghai, China. This workshop is considered to be an important event for the LarKC developers and users located in Asia.

- The 3rd LarKC early adopter tutorial was co-located with the ESWC 2010.
- The 4th LarKC early adopter workshop was designed to be an event for gathering of Chinese LarKC developers and users to exchange their ideas.
- The workshop of IWEST2010 (`http://oeg-lia3.dia.fi.upm.es/iwest2010`) was organized to prompt the cooperation between the SEALS project and LarKC.
- The 2010 International Workshop on Web-scale Knowledge Representation, Retrieval, and Reasoning (WebKR3 2010), was held on August 31, 2010, in Toronto, Canada. The workshop WebKR3 2010 was organized to prompt the cooperation between the semantic web community and the Web Intelligence community (`http://wi-consortium.org`) on Web scale data processing and reasoning.
- The ISWC2011 Joint Workshop on Knowledge Evolution and Ontology Dynamics, October 24, 2011, Bonn, Germany. This workshop aims at the research issues of ontology dynamics and evolution.
- The 2011 International Workshop on Web-scale Knowledge Representation, Retrieval, and Reasoning (WebKR3 2011), August 22, 2011, Lyon, France. This event was designed to prompt the cooperation between the community of the Semantic Web and the community of Web Intelligence.

4.3 LarKC Workflow Challenge 2011

The LarKC consortium invites the developers to take part in the 2011 LarKC Workflow Challenge. Learning to use the LarKC platform is a great skill for getting ahead of the semantic mash-up programming revolution. The LarKC Consortium offers the prizes for building fun applications as the participants learn. The challenge is: Use the Large Knowledge Collider to build the most innovative, surprising or sophisticated workflow. The challenge is very open, and meant to maximise the participant's own creativity. The only requirement is that the participants use the Large Knowledge Collider to build an interesting workflow. All the rest is up to the participants.

4.3.1 What to build?

The challenge consists in building a LarKC workflow, using either existing plugins from the LarKC marketplace, or the participants can of course write their own plugins for LarKC; plugins may execute locally or remotely; the participants can use any dataset of their own choosing, either those from the LarKC consortium, anything the participants can find on the Web of Data (either dump or SPARQL endpoint), or any other data the participants have access to; the data the participants use can be in any format, either already in RDF or still to be triplified, and can come from any data source the participants like, either static or dynamic, commercial or private or public.

The workflow can perform any task of the participants' choosing. To give the participants some ideas, we have seen and built great workflows that

- combine and interpret data streams from social networks,
- integrate data from traffic flows, weather reports, city maps and POIs,

- analyse collections of research papers hunting for as yet unknown genes implicated in cancer,
- give location-aware info to city-travellers,
- integrate massive amounts of heterogeneous biomedical data, but also
- very simple workflows that link personal calendar information with travel information.

4.3.2 How to participate?

To participate, a statement of interest was to be sent to
`larkc-challenge@lists.sourceforge.net`
before 20th July, briefly describing the plans. Before 1 September 2011 the participants were asked to submit the following:

- the workflow description in RDF, executable by LarKC V2.5;
- the required plugins (built as jar or larkc files);
- any local datasets that are required to run your workflow, if not included in the plugins;
- a brief document describing your workflow (including a how-to-start guide).

These can be submitted by email to `larkc-challenge@lists.sourceforge.net`, or sent as a pointer to a download location if a submission is too large for email.

4.3.3 Legalities

The Jury panel consists of Michael Witbrock (Cycorp Europe, chair), Tony Lee (SaltLux), and Alexey Cheptsov (HLRS). People working in the groups that form the LarKC consortium are not allowed to participate.

Submitting the workflow and plugins implies permission for the jury panel to execute the workflow. The participants are required to specify any further licensing conditions where appropriate.

5 Plan for Exploiting LarKC after LarKC

5.1 Web of Data Interpreter, A LarKC Spin-Off

Web of Data Interpreter (WoDI) is a recently launched spin-off company from the LarKC project, currently located in Innsbruck, Austria. The targeted development segment of WoDI is the implementation of intelligent tools and methods for accessing, reasoning and consuming linked data. The main areas of WoDI innovation are scalable reasoning with rules and streams of data, and pluggable and modular architectures for semantic systems based on LarKC results and technology.

5.1.1 Team

The founders of WoDI have many years of experience with semantic technologies and software engineering, and were actively involved in the LarKC project as researchers and developers.

5.1.2 Products

The methods and algorithms developed in LarKC for reasoning with very large datasets, including both static data and data streams, are the starting point in the development of WoDIs business intelligence solutions for areas such as eTourism and eCommerce. On the one hand, WoDI is exploiting the results from research in stream reasoning for building scalable solutions for logical reasoning in real time on very large and noisy data streams. Processing and reasoning over data streams is a challenging task that is gaining considerable attention given the increasing number of applications that publish online streams of data such as Twitter, or Facebook, and the multitude of sensors and aggregators that generate streams of information from the physical world. WoDI is building solutions that process and reason over data streams enabling intelligent integration and usage of relevant information. Inspired by LarKC results in the area of stream reasoning, WoDI aims at building a stream reasoning engine (called WoDI Turbine) that will meet the requirements and expectations of commercial, real-world environments. As a baseline, WoDI will leverage the architecture and design principles behind the LarKC platform, and adapt and customize the modular, pluggable, and scalable LarKC 2.5 platform for its integrated business solutions.

The data processing and reasoning tools of WoDI will not only be realized with data streams in mind, but also for working with static, interlinked data that is available as part of the Linking Open Data cloud. The importance for novel business intelligence solutions based on reasoning over streams and static data channels in combination with various other data sources becomes especially important in todays dynamic, information-driven world. Website

The website of the spin-off is localted at http://www.wodi.eu.

5.1.3 Contact

Web of Data Interpreter
Soho II

Grabenweg 68
6020 Innsbruck, Austria

E-mail: info@ wodi.eu
Phone: +43 664 88549109

5.2 Keeping LarKC Alive

The effort for keeping LarKC alive after the LarKC project is planned based on the following two kinds of activities:

- social tasks (organising events, maintaining website, social media, involving partner organisations, obtaining sponsors, etc)
- technical tasks (keeping track of bug-reports, fixing bugs, or finding, people that can fix them, maintaining source-code repository, maintaining documentation)

The total effort is estimated at 1 fte, however, not necessarily 1 person.

We are aiming at the following communities for the activities (In order of likelihood):

- life-science
- semantic mash-ups (both gov data and in the mobile domain)
- Semantic Web university teaching
- parallisation

of couse the decision will be: (in the longer run) should we do a single community or many?

The following are some examples of how to grow to a user-base

- MediaWiki: This is a "single home-run" scenario (for ?MediaWiki this was ?WikiPedia). That's not a plausible scenario for LarKC, certainly not in the short term.
- Sesame: their success is due to two factors: first to market, and ease of use. Both of these factors can be exploited for LarKC as well.
- TomCat, Drupal: For both of these, besides ease of use, an important factor was that they were free and came with a very lightweight license.

Thus, for LarKC we are in a reasonable position if we exploit our position on: i) first to market, ii) ease of use, and iii) very open licensing.

Dependence on commercial software A possible barrier to the open source future of LarKC is the dependence on OWLIM as its data-layer. Ideally, we would remove this dependence. Estimate is that this will take upwards of a month of programming.

Legal Entity It might be useful to set up a lightweight legal entity. Relationship between this entity and the existing partners (e.g. for ownership of IPR) is to be decided.

A solution would be to establish a working group of STI International[1]. This saves us any need for lawyers and other stuff.

Contributors We are soliciting support from organisations inside and outside the consortium to contribute (small amounts of) manpower to keep LarKC support running for at least the first year.Current commitments at 0.2fte for 1 year initially from

- UIBK
- VUA
- JSI
- WICI
- SoftGress

[1] http://www.sti2.org/organisation/working-groups

6 Concluding Remarks

In this document, we have analyzed various communication channels for the community-building and cross-fertilization efforts of the LarKC project. The LarKC website, the LarKC mailing lists, the LarKC Wiki, the LarKC Blog, and the LarKC Tweets have served as infrastructures for internal and external communication and have enabled knowledge sharing among internal and external members. We have reported various activities of the community-building efforts and cross-fertilization for LarKC, which include the activities related-to the cooperation with other European projects, such as SEALS and OpenPhacts, and various events that were organized or attended by LarKC consortium members for community-building and cross-fertilization efforts, including several workshops, such as WebKR3 2011. In this document, we have also presented the plan for exploiting LarKC after the LarKC project is finished.