LarKC

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

FP7 – 215535

D6.4 – 1st periodic report on data and performances

Coordinator: Emanuele Della Valle

With contributions from: Daniele Dell'Aglio, Irene Celino

Quality Assessor: Georgina Gallizo

Quality Controller: Emanuele Della Valle

Document Identifier:	LarKC/2008/D6.4/V1.0
Class Deliverable:	LarKC EU-IST-2008-215535
Version:	Version 1.0
Date:	May 28th, 2009
State:	Final
Distribution:	Public

EXECUTIVE SUMMARY

This document is concerned with the periodic report on data and performances. It provides a regular report on the volume of data acquired and the performance of the LarKC platform. This report therefore provides measuring and quantifying progress and impact of LarKC project. By this template, LarKC consortium and other interest groups can recognize easily the status and achievement of this project on what the Urban Computing use case is concerned.

For reporting on the data this report - which follows the template provided in D6.2 "Templates of periodic report on data and performances" - includes the data source metadata, the "semantics" of the data source, and the data source format, in order to describe data characteristics, usages, size and so on.

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**	6.4	**Title**	1st periodic report on data and performances
Work Package	**Number**	6	**Title**	Urban Computing

Date of Delivery	**Contractual**	M14	**Actual**	M14
Status	final		final ■	
Nature	prototype □ report ■ dissemination □			
Dissemination level	public ■ consortium □			

Authors (Partner)	Daniele Dell'Aglio (Cefriel), Irene Celino (Cefriel)			
Responsible Author	**Name**	Emanuele Della Valle	**E-mail**	emanuele.dellavalle@cefriel.it
	Partner	Cefriel	**Phone**	+39 (02) 23954-324

Abstract (for dissemination)	This report is about the activities of the first period of the project related to the data collection activities in the field of Urban Computing.
Keywords	data sets, use case, urban computing, measure, periodic report

Version Log			
Issue Date	**Rev. No.**	**Author**	**Change**
May 6, 2009	0.1	Daniele	First draft of the document
May 14, 2009	0.2	Irene	Finalization of the document
May 26, 2009	0.3	Emanuele	Updated plan for performance measures
May 28, 2009	0.4	Irene	Document review on the basis of the QA feedbacks

PROJECT CONSORTIUM INFORMATION

Participant's name	Partner	Contact
Semantic Technology Institute Innsbruck, Universitaet Innsbruck		Prof. Dr. 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.		Dr. Lael Schooler Max-Planck-Institut für Bildungsforschung Berlin, Germany Email: schooler@mpib-berlin.mpg.de
Ontotext Lab, Sirma Group Corp		Atanas Kiryakov, Ontotext Lab, Sofia, Bulgaria Email: atanas.kiryakov@sirma.bg
SALTLUX INC.		Tony Lee, SALTLUX INC, Seoul, Korea, Email: tony@saltlux.com
SIEMENS AKTIENGESELLSCHAFT		Dr. Volker Tresp, SIEMENS AKTIENGESELLSCHAFT, Muenchen, Germany, E-mail: volker.tresp@siemens.com
THE UNIVERSITY OF SHEFFIELD		Prof. Dr. Hamish Cunningham, THE UNIVERSITY OF SHEFFIELD Sheffield, UK, Email: h.cunningham@dcs.shef.ac.uk

VRIJE UNIVERSITEIT AMSTERDAM		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		Prof. Dr. Ning Zhong, THE INTERNATIONAL WIC INSTITUTE, Mabeshi, Japan, Email: zhong@maebashi-it.ac.jp
INTERNATIONAL AGENCY FOR RESEARCH ON CANCER		Dr. Paul Brennan, INTERNATIONAL AGENCY FOR RESEARCH ON CANCER, Lyon, France, Email: brennan@iarc.fr

TABLE OF CONTENTS

1. Introduction

This deliverable aims to provide an update on the activities carried on within WP6 for what regards the collection, gathering and analysis of available data sources that can be useful in setting up concrete Urban Computing scenarios for demonstrating LarKC technologies.

In this first report, we provide the description of the data sets we collected in the first part of the project and on which we are experimenting the first outcomes of the LarKC platform. The section about performance measurements of this deliverable was set up, but this document does not contain any figure about performances, because a stable version of the platform has not been released yet.

This document therefore contains some tables to describe the collected data; each of the table is compliant with the template we provided in deliverable D6.2 “Templates of periodic report on data and performances” [1], which was designed to contain statistical and quantitative and qualitative indicators for measuring and quantifying progress and impact of LarKC project in what the Urban Computing use case is concerned. By using this template, LarKC consortium and other interest groups can recognize easily the status and achievement of this project. This template of the periodic report on data and performances plays a role not just measuring and evaluating this project but also providing and communicating the whole LarKC deliverables to our final users.

Using the same templates defined in deliverable D6.2, we will continue to periodically report on data and performances to show the progress of the LarKC project from the point of view of the Urban Computing use case at M18, M26, M33, and M42. (The corresponding deliverables are D6.6, D6.7, D6.8, and D6.11 respectively.)

2. Periodic report on Data

2.1. Data retrieved from AMA (Agenzia Mobilità e Ambiente) Milano

Connecting and registering to the AMA Web site[1], it is possible to download some data sets containing information about Milano and its hinterland. The format of this data is the ESRI shapefile, compatible with some GIS systems[2]. In this section we will present the data sets that we got from AMA.

Data Source: Graph of Milano			
Report ID			
Section 1	**Data source metadata**		
Name	Graph of Milano		
Producer/Owner	Agenzia Mobilità e Ambiente Srl (http://www.ama-mi.it)		
Description	Data about the road network of Milano and some municipalities (the hinterland) around it. It contains: 1. directed graph of the road network 2. turning prohibitions 3. outflow BVR curve values		
Namespace/Web Address	http://81.208.25.93/documenti/grafo.pdf (in Italian)		
Availability	A registration is required to access the data (and the data cannot be redistributed)		
Download/Upload/Acquisition date	October 20, 2008		
Version	0.4 (April 15, 2008)		
Physical size	5,64 MB		
Nature of data type	Static data		
Quality of the data source	Good		
Section 2	**"semantics" of the data source**		
Typology of data	Topology		
Geographic coverage of data	City (Milano and hinterland)		
Applied systems	GIS, Geoinformatics		
Existence of schema/ontology	http://wiki.larkc.eu/LarkcProject/WP6/WorkInProgress/AMAData (see remarks)		
Existing links with other data-sources	Vertexes identifier allow to link this data source to Statistical traffic data source (see below: "Statistical traffic data" dataset)		
Possible linkage to other data-sources	Coordinates to link this data to external data sources		
Scale of data	The graph contains about 14.000 nodes and 28.000 links and their characteristics		
Section 3	**Data source format**		
Format of data	GIS specific format		
Generation method	Not available		
Support query language	See remarks	**Total no. of statements**	About 300.000 (see remarks)
Support triple type	See remarks		
No. of explicit statements	About 300.000 (see remarks)		
Noise, Uncertainty and inconsistence of data	The graph is not strongly connected.		
Remarks			
	This data has been converted in RDF. See D6.3 [2], Section 2.1, to get more information about the conversion process.		

[1] http://www.ama-mi.it/
[2] More information available here: http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?id=2729&pid=2727&topicname=Shapefile_file_extensions

Data Source: O/D matrices of Milano			
Report ID			
Section 1	**Data source metadata**		
Name	O/D matrices of Milano (where O/D means origin/destination)		
Producer/Owner	Agenzia Mobilità e Ambiente Srl (http://www.ama-mi.it)		
Description	This data source is composed by three O/D matrices that contain information about the movement of Milano people (by car or motorcycle) inside the city in different time intervals (morning, afternoon and evening). Milano (and its hinterland) has been partitioned in (about) 600 zones: for every couple of them there is the number of vehicle that starts from the first zone and arrive in the second one.		
Namespace/Web Address	http://81.208.25.93/documenti/matrici.pdf (in Italian)		
Availability	A registration is required to access the data (and the data cannot be redistributed)		
Download/Upload/Acquisition date	October 20, 2008		
Version	0.4 (April 15, 2008)		
Physical size	7,95 MB		
Nature of data type	Static data		
Quality of the data source	Good		
Section 2	**"semantics" of the data source**		
Typology of data	Statistical information; Traffic information		
Geographic coverage of data	City (Milano and hinterland)		
Applied systems	GIS, Geoinformatics, Traffic engineering		
Existence of schema/ontology	http://wiki.larkc.eu/LarkcProject/WP6/WorkInProgress/AMAData (see remarks)		
Existing links with other data-sources	No		
Possible linkage to other data-sources	For every zone is defined the list of coordinates that define its perimeter. This geographical information could be used to link this data source with others		
Scale of data	See below		
Section 3	**Data source format**		
Format of data	GIS specific format		
Generation method	Hand-made (interviews)		
Support query language	SPARQL (see remarks)	**Total no. of statements**	About 700.000 (see remarks)
Support triple type	See remarks		
No. of explicit statements	About 700.000 (see remarks)		
Noise, Uncertainty and inconsistence of data	-		
Remarks			
	This data has been converted in RDF. See D6.3 [2], Section 2.1, to get more information about the conversion process.		

Data Source: Statistical traffic data			
Report ID			
Section 1	**Data source metadata**		
Name	Statistical traffic data		
Producer/Owner	Agenzia Mobilità e Ambiente Srl (http://www.ama-mi.it)		
Description	This data source contains the number of vehicles that passed through some section of roads. The data is aggregated and numbers are related to time ranges (annual, from 2001 to 2006)		
Namespace/Web Address	http://81.208.25.93/documenti/Dati_Traffico.pdf (in Italian)		
Availability	A registration is required to access the data (and it could not be redistributed)		
Download/Upload/Acquisition date	October 20, 2008		
Version	0.2 (May 07, 2008)		
Physical size	46 KB		
Nature of data type	Static data		
Quality of the data source	Good		
Section 2	**"semantics" of the data source**		
Typology of data	Traffic information; Statistical information		
Geographic coverage of data	City (Milano)		
Applied systems	Geoinformatics, Traffic engineering		
Existence of schema/ontology	http://wiki.larkc.eu/LarkcProject/WP6/WorkInProgress/AMAData (see remarks)		
Existing links with other data-sources	Road sections are defined with start and ending node identifiers that allow to link this data to Graph of Milano		
Possible linkage to other data-sources	-		
Scale of data	See below		
Section 3	**Data source format**		
Format of data	Excel document		
Generation method	Original data has been collected with both manual and automatic surveys. The generation of aggregate data is done with automatic processes.		
Support query language	SPARQL (see remarks)	**Total no. of statements**	About 2.500 (see remarks)
Support triple type	See remarks		
No. of explicit statements	About 2.500 (see remarks)		
Noise, Uncertainty and inconsistence of data	-		
Remarks			
	This data has been converted in RDF. See D6.3 [2], Section 2.1, to get more information about the conversion process.		

Data Source: O/D matrices (for goods) of Milano			
Report ID			
Section 1	**Data source metadata**		
Name	O/D matrices (for goods) of Milano		
Producer/Owner	Agenzia Mobilità e Ambiente Srl (http://www.ama-mi.it)		
Description	These O/D matrices contains information about the movements of goods in Milano (partitioned in about 130 zones)		
Namespace/Web Address	http://81.208.25.93/documenti/MatrMerci.pdf (in Italian)		
Availability	A registration is required to access the data (and it could not be redistributed)		
Download/Upload/Acquisition date	October 20, 2008		
Version	0.1 (October 11, 2007)		
Physical size	896 KB		
Nature of data type	Static data		
Quality of the data source	Good		
Section 2	**"semantics" of the data source**		
Typology of data	Statistical information; Traffic information		
Geographic coverage of data	City (Milano and hinterland)		
Applied systems	GIS, Geoinformatics, Traffic engineering		
Existence of schema/ontology	See remarks		
Existing links with other data-sources	No		
Possible linkage to other data-sources	For every zone is defined the list of coordinates that define its perimeter. This geographical information could be used to link this data source with others		
Scale of data	See remarks		
Section 3	**Data source format**		
Format of data	GIS specific format		
Generation method	Man-made (interviews) and automatic		
Support query language	See remarks	**Total no. of statements**	See remarks
Support triple type	See remarks		
No. of explicit statements	See remarks		
Noise, Uncertainty and inconsistence of data	-		
Remarks			
	The schema of this data set is the same of O/D matrices of Milano. We will convert this data set in RDF when we will start to use this information for traffic prediction (and so on).		

2.2. Data from the Web

The biggest archive of data sources that we can found is the Web. We consider two main groups:

- **Linking Open Data** (*LOD*): closely related to Semantic Web activities, this community project is trying to build a network of open data sets. As the name suggests, the objective of LOD is to connect different sets of publicly accessible data containing links to other data sources. This project is growing more and more and we can find a lot of useful data for the Urban Computing scenarios. Actually we started to consider two datasets: GeoNames and DBPedia.
- **Other data sources**: there are also a lot of Web data sources with data that can be useful for our purposes. For example in Web 2.0 there are a lot of sites there the contents are generated (totally or partially) by users. Often this kind of Web sites offers also an interface accessible from applications, like a REST service. This fact allows the creation of new applications for example mixing the contents of different data sets (mash-ups). We choose some data sources containing information that could allow the linkage to the other data sets we are working on: Eventful, Last.fm and Yahoo! Upcoming.

In the following we present these data sources describing their characteristics.

Data Source: DBpedia			
Report ID			
Section 1	**Data source metadata**		
Name	DBpedia		
Producer/Owner	Community project (more info at: http://wiki.dbpedia.org/Team)		
Description	This project try to extract the structured information contained in Wikipedia and make it available in Web of data.		
Namespace/Web Address	http://dbpedia.org (Web site) http://dbpedia.org/sparql (SPARQL endpoint) http://dbpedia.org/snorql/ (SNORQL interface)		
Availability	GNU FDL		
Download/Upload/Acquisition date	May 7, 2009		
Version	3.2		
Physical size	1,3 GB		
Nature of data type	Dynamic		
Quality of the data source	Good		
Section 2	**"semantics" of the data source**		
Typology of data	Encyclopaedia		
Geographic coverage of data	World		
Applied systems	Semantic Web, Web applications		
Existence of schema/ontology	http://downloads.dbpedia.org/3.2/en/dbpedia-ontology.owl (http://www4.wiwiss.fu-berlin.de/dbpedia/dev/ontology.htm)		
Existing links with other data-sources	It is part of the LOD network (more info at http://linkeddata.org/)		
Possible linkage to other data-sources	Coordinates to link this data to other datasets containing geo-location information		
Scale of data	http://wiki.dbpedia.org/Ontology		
Section 3	**Data source format**		
Format of data	RDF		
Generation method	Automatic		
Support query language	SPARQL	**Total no. of statements**	More than 250 million
Support triple type	N-Triples		
No. of explicit statements	More than 250 million		
Noise, Uncertainty and inconsistence of data	Same data "precision" as in Wikipedia		
Remarks			
	-		

Data Source: GeoNames			
Report ID			
Section 1	**Data source metadata**		
Name	GeoNames		
Producer/Owner	Marc Wick (Founder)		
Description	It is a geographical archive containing relevant places (countries, cities, hotels, monuments and so on) with their names and coordinates to locate them. It also contains a database with postal codes of different countries and cities.		
Namespace/Web Address	http://www.geonames.org		
Availability	Creative Commons 3.0 BY		
Download/Upload/Acquisition date	May 7, 2009		
Version	2.0		
Physical size	1 GB		
Nature of data type	Dynamic		
Quality of the data source	Good		
Section 2	**"semantics" of the data source**		
Typology of data	Geo-locations		
Geographic coverage of data	World		
Applied systems	Geoinformatics, Web application, Semantic Web		
Existence of schema/ontology	OWL full: http://www.geonames.org/ontology/ontology_v2.0_Full.rdf OWL lite: http://www.geonames.org/ontology/ontology_v2.0_Lite.rdf		
Existing links with other data-sources	It is part of the LOD network (more info at http://linkeddata.org/)		
Possible linkage to other data-sources	Coordinates to link this data to other datasets containing geo-location information		
Scale of data	It contains over eight million geographical names and consists of 6.5 million unique features whereof 2.2 million populated places and 1.8 million alternate names. All features are categorized into one out of nine feature classes and further subcategorized into one out of 645 feature codes. (from: http://www.geonames.org/about.html)		
Section 3	**Data source format**		
Format of data	• XML, JSON, RDF, CSV, TXT, RSS, KML (REST services http://www.geonames.org/export/ws-overview.html) • RDF dump (http://www.geonames.org/ontology/)		
Generation method	Data is generated from different sources. In addition users can edit it (to add new data or to fix errors).		
Support query language	Via REST service (see above)	**Total no. of statements**	About 100 million
Support triple type	RDF		
No. of explicit statements	About 100 million		
Noise, Uncertainty and inconsistence of data	-		
Remarks			
	-		

Data Source: Eventful			
Report ID			
Section 1	**Data source metadata**		
Name	Eventful		
Producer/Owner	Eventful, Inc.		
Description	Eventful is a Web 2.0 site where users can insert information about events: the kind of event (music, sport…), its location, its start time and so on.		
Namespace/Web Address	http://www.eventful.com (Web site) http://api.eventful.com/rest/ (REST entry point) http://api.eventful.com/ (API reference)		
Availability	The data is publicly accessible in the Web site. The access to the REST service requires an API key (that can be obtained with a registration). In addition, the logo of Eventful should be inserted in external web pages that use its contents (more information at: http://api.eventful.com/terms)		
Download/Upload/Acquisition date	May 7, 2009		
Version	Not available		
Physical size	Not available (it's a REST service)		
Nature of data type	Dynamic		
Quality of the data source	The data is inserted by users, so the update frequency depends by their participation. The quality of data changes (some events are described very well and other ones are incomplete or wrong)		
Section 2	**"semantics" of the data source**		
Typology of data	Events, Venues		
Geographic coverage of data	World		
Applied systems	Web site users, Web 2.0 applications		
Existence of schema/ontology	Custom XML schema (described at: http://api.eventful.com/tools/feeds)		
Existing links with other data-sources	Maybe – answers about events, venues and performers can contain links to other Web sites (for example the Official Web site of a singer). It means that there could be a link to other data sources.		
Possible linkage to other data-sources			
Scale of data	About 130.000 events per week		
Section 3	**Data source format**		
Format of data	XML, RSS, iCal (REST service)		
Generation method	Man-made (user generated)		
Support query language	See remarks	**Total no. of statements**	Not available (it's not possible to obtain the whole data set)
Support triple type	See remarks		
No. of explicit statements	Not available		
Noise, Uncertainty and inconsistence of data	1. The same event could be inserted two or more times (for example: http://eventful.com/encino/events/old-time-music-jam-/E0-001-020726420-2 and http://eventful.com/encino/events/old-time-music-jam-/E0-001-018341075-6) 2. The event could be not sure (for example matches in NBA playoffs http://en.wikipedia.org/wiki/NBA_Playoffs) 3. The data can contain wrong or unclear information		
Remarks			
	The natively data is not in a RDF format. We can perform a GRDDL transformation to obtain RDF data.		

Data Source: Last.fm			
Report ID			
Section 1	**Data source metadata**		
Name	Last.fm		
Producer/Owner	Last.fm Ltd (CBS Interactive)		
Description	Last.fm is one of the most popular Web sites. It offers a user-personalized Web radio, a recommendation system, a search engine to find music events in the World…		
Namespace/Web Address	http://last.fm (Web site) http://ws.audioscrobbler.com/2.0/ (REST entry point) http://www.last.fm/api (API reference)		
Availability	Data could be used, published and distributed in the original form and in derivate works for non-commercial purposes and citing last.fm name and the URL of the Web site. (more details at: http://www.last.fm/api/tos - section 4)		
Download/Upload/Acquisition date	May 7, 2009		
Version	Not available		
Physical size	Not available (it's a Web service)		
Nature of data type	The data is inserted by maintainers, labels/musicians (songs) and users (tags, pics…).		
Quality of the data source	Probably one of the most complete archive about music available on the Web.		
Section 2	**"semantics" of the data source**		
Typology of data	Music (artist and song information, venues, events)		
Geographic coverage of data	World		
Applied systems	Web site users, Web 2.0 applications		
Existence of schema/ontology	Custom XML schema (described in API reference: http://www.last.fm/api)		
Existing links with other data-sources	No		
Possible linkage to other data-sources	It could be possible to link this data to other data sources (for example using geo-location of venues)		
Scale of data	Not available		
Section 3	**Data source format**		
Format of data	XML, RSS, JSON, iCal (REST service)		
Generation method	The method are both manually (a review inserted by an user) and automatic (a recommendation)		
Support query language	See remarks	**Total no. of statements**	Not available (it's not possible to obtain the whole data set)
Support triple type	See remarks		
No. of explicit statements	Not available		
Noise, Uncertainty and inconsistence of data	The same event (or venue) could be inserted two or more times		
Remarks			
	The natively data is not in a RDF format. We can perform a GRDDL transformation to obtain RDF data.		

<table>
<tr><td colspan="4">Data Source: Yahoo! Upcoming</td></tr>
<tr><td>Report ID</td><td colspan="3"></td></tr>
<tr><td>Section 1</td><td colspan="3">Data source metadata</td></tr>
<tr><td>Name</td><td colspan="3">Upcoming</td></tr>
<tr><td>Producer/Owner</td><td colspan="3">Yahoo!</td></tr>
<tr><td>Description</td><td colspan="3">Upcoming is a Web site similar to Eventful: users could register themselves and insert events and related information (kind, location…). Then they could manage a calendar subscribing the events that they want to join and share these information with friends.</td></tr>
<tr><td>Namespace/Web Address</td><td colspan="3">http://upcoming.yahoo.com (Web site)
http://upcoming.yahooapis.com/services/rest/ (REST entry point)
http://upcoming.yahoo.com/services/api/ (API reference)</td></tr>
<tr><td>Availability</td><td colspan="3">The access to the REST service for non-commercial use requires an API key that can be obtained with a registration.</td></tr>
<tr><td>Download/Upload/Acquisition date</td><td colspan="3">May 7, 2009</td></tr>
<tr><td>Version</td><td colspan="3">2.0</td></tr>
<tr><td>Physical size</td><td colspan="3">Not available (it’s a REST service)</td></tr>
<tr><td>Nature of data type</td><td colspan="3">User-generated contents</td></tr>
<tr><td>Quality of the data source</td><td colspan="3">The data is inserted by users, so the update frequency depends by their participation. The quality of data changes (some events are described very well and other ones are incomplete or wrong)</td></tr>
<tr><td>Section 2</td><td colspan="3">“semantics” of the data source</td></tr>
<tr><td>Typology of data</td><td colspan="3">Events</td></tr>
<tr><td>Geographic coverage of data</td><td colspan="3">World</td></tr>
<tr><td>Applied systems</td><td colspan="3">Web site users, Web 2.0 applications</td></tr>
<tr><td>Existence of schema/ontology</td><td colspan="3">Custom XML schema
(described in API reference: http://upcoming.yahoo.com/services/api/)</td></tr>
<tr><td>Existing links with other data-sources</td><td colspan="3" rowspan="2">Maybe – answers about events, venues, users, etc. can contain links to other Web sites (for example the Official Web site of a singer). It means that there could be a link to other data sources.</td></tr>
<tr><td>Possible linkage to other data-sources</td></tr>
<tr><td>Scale of data</td><td colspan="3"></td></tr>
<tr><td>Section 3</td><td colspan="3">Data source format</td></tr>
<tr><td>Format of data</td><td colspan="3">XML, JSON (REST service)</td></tr>
<tr><td>Generation method</td><td colspan="3">Man-made (by users)</td></tr>
<tr><td>Support query language</td><td>See remarks</td><td>Total no. of statements</td><td>Not available (it’s not possible to obtain the whole data set)</td></tr>
<tr><td>Support triple type</td><td colspan="3">See remarks</td></tr>
<tr><td>No. of explicit statements</td><td colspan="3">Not available</td></tr>
<tr><td>Noise, Uncertainty and inconsistence of data</td><td colspan="3">1. The same event could be inserted two or more times
2. The event could be not sure (for example matches in NBA playoffs)
3. The data can contain wrong or unclear information</td></tr>
<tr><td>Remarks</td><td colspan="3"></td></tr>
<tr><td></td><td colspan="3">The natively data is not in a RDF format. We can perform a GRDDL transformation to obtain RDF data.</td></tr>
</table>

3. Periodic report on Performances

Due to the fact that the first release of the LarKC platform will be available only at the end of May, we prepared the environment to perform system performances analyses. Our evaluation strategy is to compare LarKC performances in solving a Urban Computing problem with a "term of comparison", a software product which achieves the same solution without using LarKC technologies. More details are available in Section 2.3 of deliverable D6.3 [2]. The results will be made available in D6.5.

.

4. Conclusions

In this deliverable, we provided an update on the activities carried on within WP6 for what regards the collection of data sources. We explained what data we are currently taking into consideration to fulfil some requirements in Urban Computing scenarios.

The next version of this deliverable (D6.6 – Second periodic report on data and performances) will also contain the first results about the performances of LarKC on what the Urban Computing use case is concerned.

5. References

[1] Kono Kim, Irene Celino, Emanuele Della Valle, Daniele Dell'Aglio, Yi Huang, Werner Hauptmann – ***Deliverable D6.2 "Templates of periodic report on data and performances"***, LarKC project deliverable, December 2008, available from http://www.larkc.eu/deliverables/

[2] Emanuele Della Valle, Daniele Dell'Aglio, Irene Celino – ***Deliverable D6.3 "Urban Computing environment specification"***, LarKC project deliverable, May 2009, available from http://www.larkc.eu/deliverables/