



LarKC

The Large Knowledge Collider

a platform for large scale integrated reasoning and Web-search

FP7 – 215535

D1.3.2 Improved LarKC Plugin Annotation Language and Methodology

Coordinator: Ioan Toma (STI Innsbruck)

With contributions from:

Quality Assessor: Irene Celino (CEFRIEL)

Quality Controller: Reto Krummenacker (STI Innsbruck)

Document Identifier:	LarKC/2008/D1.3.2/V1.0
Class Deliverable:	LarKC EU-IST-2008-215535
Version:	version 1.0.0
Date:	January 2, 2011
State:	final
Distribution:	public



EXECUTIVE SUMMARY

This document provides an improved annotation ontology for describing LarKC plug-ins. The focus is on the QoS aspects of plugins, the set of non-functional properties covered in the ontology being larger, much clearer and better formalized than in the initial version - D1.3.1 "Initial Initial Plug-in AnnotationLanguage". Of equal importance to describing plug-ins is to describe datasets that are used by LarKC plug-ins to reason upon them. This document also contains the void descriptions of two important LarKC datasets, namely Linked Life Data and FactForge.



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	1.3.2	Title	Improved LarKC Plugin Annotation Language and Methodology
Work Package Evaluation	Number	1	Title	Conceptual Framework

Date of Delivery	Contractual	M33	Actual	30-Dec-10
Status	version 1.0.0		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)	STI Innsbruck			
Resp. Author	Ioan Toma		E-mail	ioan.toma@sti2.at
	Partner	STI Innsbruck	Phone	+43 512 507 6485

Abstract (for dissemination)	This document provides an improved annotation ontology for describing LarKC plug-ins. The focus is on the QoS aspects of plugins, the set of non-functional properties covered in the ontology being larger, much clear and better formalized than in the initial version - D1.3.1 "Initial Initial Plug-in AnnotationLanguage". Of equal importance to describing plug-ins is to describe datasets that are used by LarKC plug-ins to reason upon them. This document also contains the void descriptions of two important LarKC datasets, namely Linked Life Data and FactForge.
Keywords	Plugin Annotation, Non-functional properties ontology, Dataset, void



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 Email: 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 RE- SPONSABILITA LIMITATA Milano, Italy Email: emanuele.dellavalle@cefriel.it
CYCORP, RAZISKOVANJE IN EKSPERI- MENTALNI RAZVOJ D.O.O.		Michael Witbrock CYCORP, RAZISKOVANJE IN EKSPERI- MENTALNI 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 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		Dr. Volker Tresp SIEMENS AKTIENGESELLSCHAFT Muenchen, Germany Email: 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 INSTI- TUTE, BEIJING UNIVERSITY OF TECHNOLOGY		Prof. Dr. Ning Zhong, THE INTERNATIONAL WIC INSTITUTE Mabeshi, Japan Email: zhong@maebashi-it.ac.jp
INTERNATIONAL AGENCY FOR RE- SEARCH ON CANCER		Dr. Paul Brennan, INTERNATIONAL AGENCY FOR RE- SEARCH ON CANCER Lyon, France Email: brennan@iarc.fr
INFORMATION RETRIEVAL FACILITY		Dr. John Tait, Dr. Paul Brennan, INFORMATION RETRIEVAL FACILITY Vienna, Austria Email: john.tait@ir-facility.org





TECHNICAL UNIVERSITY OF CLUJ-NAPOCA http://www.utcluj.ro/	 The logo of the Technical University of Cluj-Napoca, consisting of a stylized 'U' and 'T' in red and grey.	Prof. Dr. Eng. Sergiu Nedevschi TECHNICAL UNIVERSITY OF CLUJ-NAPOCA Cluj-Napoca, Romania E-mail: sergiu.nedevschi@cs.utcluj.ro
SOFTGRESS S.R.L. http://www.softgress.com/	 The logo for Softgress, featuring the word 'Softgress' in white text on a blue rectangular background.	Dr. Ioan Toma SOFTGRESS S.R.L. Cluj-Napoca, Romania E-mail: ioan.toma@softgress.com



TABLE OF CONTENTS

LIST OF FIGURES	7
1 INTRODUCTION	8
2 UPDATED ONTOLOGY FOR PLUG-INS NON-FUNCTIONAL PROPERTIES	9
2.1 Infrastructure	9
2.1.1 Memory	9
2.1.2 CPU	10
2.1.3 Execution Environment	11
2.1.4 Deployment Factor	11
2.2 Performance	12
2.3 Scalability	13
2.4 Accuracy	13
2.4.1 Completeness and Soundness	14
2.4.2 Precision and Recall	14
2.4.3 F-Measure, ROC Curve, AUC Curve	15
2.5 Temporal	16
2.6 Financial	16
2.7 Utilities	17
2.7.1 Cardinality	17
2.7.2 Limits	18
3 LARKC DATASETS DESCRIPTIONS	24
4 CONCLUSION	31



LIST OF FIGURES

2.1	Memory	9
2.2	CPU	10
2.3	Execution Environment	11
2.4	Deployment Factor	12
2.5	Performance	12
2.6	Scalability	13
2.7	Completeness and Soundness	14
2.8	Precision and Recall	15
2.9	F-Measure, ROC Curve, AUC Curve	15
2.10	Temporal	16
2.11	Financial	17
2.12	Cardinality	17
2.13	Limits	18



1. Introduction

The Large Knowledge Collider (LarKC) is developing a pluggable infrastructure for large scale, distributed and incomplete reasoning. The pluggable infrastructure ensures that computational components of diverse fields can be coherently integrated in order to coordinate large scale inference over distributed and heterogeneous information and resources. Such computation components referred as plug-ins play a key role in the LarKC context. They provide specific functionalities (e.g. retrieval and selection, transformation, reasoning, etc.) and can be discovered, selected and composed in more complex research processes (so-called workflows of special-purpose plug-ins). To enable the afore mentioned tasks on plug-ins explicit specifications of the functional and non-functional properties of plug-ins is required. In a earlier deliverable we identified the relevant functional and non-functional properties of plugins and proposed an initial annotation ontology to describe these properties. Taking as a starting point the initial annotation approach proposed in [1], the current deliverable extends and improves the annotation ontology for LarKC plug-ins. It includes larger and much clear and better formalized models for non-functional properties of plug-ins. Besides formal and precise descriptions of plug-ins of equal importance are the description of datasets. Formal and precise metadata about datasets enables discovery and usage of these datasets in LarKC workflows. In this context, the current deliverable also contains the void descriptions of two important LarKC datasets, namely Linked Life Data and FactForge.

The reminder of this deliverable is organized as follows. Chapter 2 takes the set of plug-in non-functional properties identified in [1], extends it with additional properties and more importantly remodels properties in a cleaner way. Examples are provided for each concept. Chapter 3 contains descriptions of two LarKC datasets, namely Linked Life Data and FactForge using void terminology. Finally, Chapter 4 summarizes the deliverable.

2. Updated ontology for plug-ins Non-functional properties

Taking as a starting point the initial annotation approach proposed in [1], we extend and improve the annotation ontology for LarKC plug-ins. The updates and improvements concern the the non-functional properties of plug-ins, while for functional properties we keep the initial approached proposed in [1]. Providing good models for non-functional properties of plug-ins, as well as of services [2], is important to enable discovery and composition. The set of non-functional properties covered in this document is larger and much clear and better formalized. Extra classes and properties were added to enable the description of temporal aspects. Additionally general purpose utility concepts and properties for limits and cardinality specification were added.

As mentioned above, the improved annotation ontology focuses on non-functional properties of plug-ins. In the following subsections we revisit the list of non-functional properties that form the ontology and present graphically the major concept. The annotation ontology is a RDFs ontology that includes non-functional properties such as infrastructure, performance, scalability, accuracy and financial aspects.

2.1 Infrastructure

The Infrastructure part of the ontology includes all notions that are used to model a particular environment within which a particular plug-in can be executed. It includes terminology for specifying the memory, CPU, execution environment, deployment factors. The following subsections detail each of concepts mentioned before.

2.1.1 Memory

A graphical representation of the memory fragment of the ontology is available in Figure 2.1. This part of the ontology is used to specify the memory requirements of a plug-in.

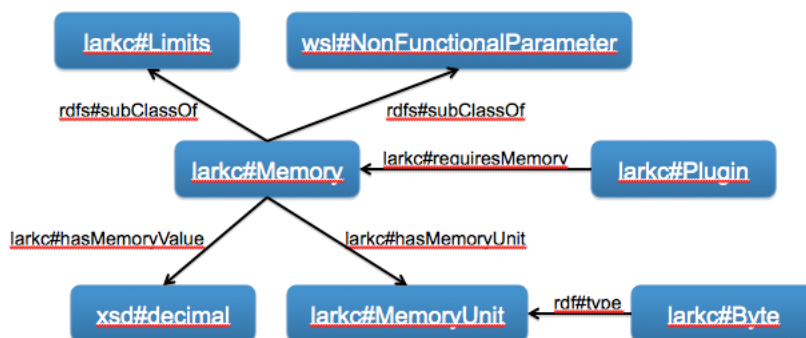


Figure 2.1: Memory

Listing 2.1.1 contains an example of how to describe the memory requirements for a LarKC plug-in.

Listing 2.1: Memory example

```
@prefix larkc: <http://larkc.eu/plugin#> .
```

```

@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ex: <http://www.example.org#> .

ex:SomeIdentifier rdf:type larkc:Identifier ;
    larkc:requiresMemory ex:Memory .

ex:Memory rdf:type larkc:Memory;
    larkc:hasMemoryUnit larkc:Bytes;
    larkc:hasMinimum "53687091200"^^integer .           //5GB
    
```

2.1.2 CPU

A graphical representation of the CPU fragment of the ontology is available in Figure 2.2. This part of the ontology is used to specify the CPU requirements of a plug-in.

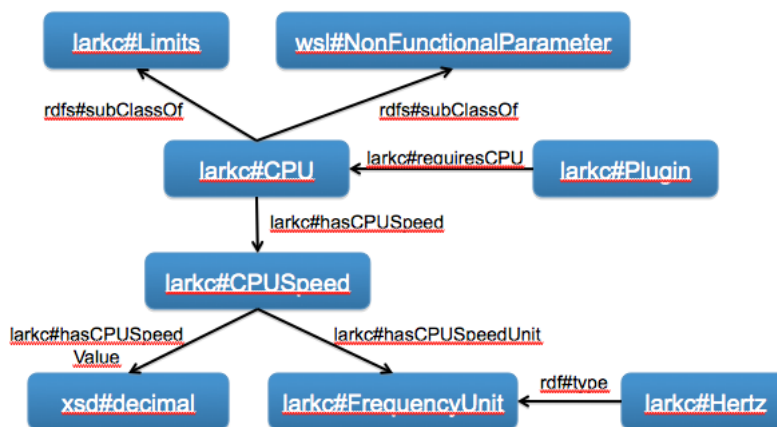


Figure 2.2: CPU

Listing 2.1.2 contains an example of how to describe the CPU requirements for a LarkC plug-in.

Listing 2.2: CPU example

```

@prefix larkc: <http://larkc.eu/plugin#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ex: <http://www.example.org#> .

ex:SomeTransformer rdf:type larkc:Transformer ;
    larkc:requiresCPU ex:CPU .

ex:CPU rdf:type larkc:CPU;
    larkc:hasCPUSpeedUnit larkc:Hertz;
    larkc:hasMinimum "30000000000"^^integer .           //3 Gigahertz
    
```

2.1.3 Execution Environment

Execution Environment refers to the concrete execution environment for a LarKC plug-in. A graphical representation of the Execution Environment fragment of the ontology is available in Figure 2.3.

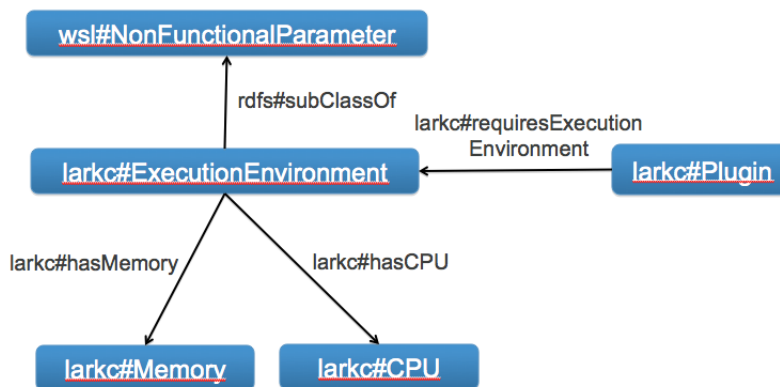


Figure 2.3: Execution Environment

Listing 2.1.3 contains an example of how to specify the required execution environment of a LarKC plug-in.

Listing 2.3: Execution Environment example

```

@prefix larkc: <http://larkc.eu/plugin#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ex: <http://www.example.org#> .

ex:SomeSelector rdf:type larkc:Selector ;
    larkc:requiresExecutionEnvironment ex:ExecutionEnvironment .

ex:ExecutionEnvironment rdf:type larkc:ExecutionEnvironment ;
    larkc:hasMemory ex:Memory ;
    larkc:hasCPU ex:CPU .

ex:Memory rdf:type larkc:Memory;
    larkc:hasMemoryUnit larkc:Bytes;
    larkc:hasMinimum "4294967296"^^integer . //4GB

ex:CPU rdf:type larkc:CPU;
    larkc:hasCPUSpeedUnit larkc:Hertz;
    larkc:hasMinimum "2500000000"^^integer . //2.5Gigahertz
    
```

2.1.4 Deployment Factor

Deployment Factor is the minimum speed of connection needed to the plug-in external world. A graphical representation of the Deployment Factor fragment of the ontology is available in Figure 2.4.

Listing 2.1.4 contains an example of how to specify the required deployment factor of a LarKC plug-in.

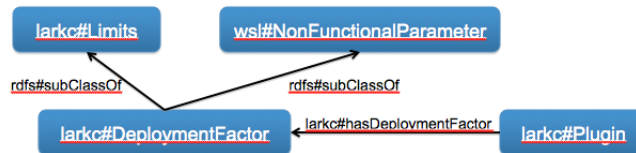


Figure 2.4: Deployment Factor

Listing 2.4: Deployment Factor example

```

@prefix larkc: <http://larkc.eu/plugin#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ex: <http://www.example.org#> .

ex:SomeIdentifier rdf:type larkc:Identifier ;
    larkc:hasDeploymentFactor ex:DeploymentFactor .

ex:DeploymentFactor rdf:type larkc:DeploymentFactor ;
    larkc:hasMinimum "1000"^^xsd:integer . //minimum speed of
        connection
    
```

2.2 Performance

Performance is related to how fast a service request can be completed and at which rate the results are generated. Central to performance is the estimated invocation time. A graphical representation of the performance fragment of the ontology is available in Figure 2.5.

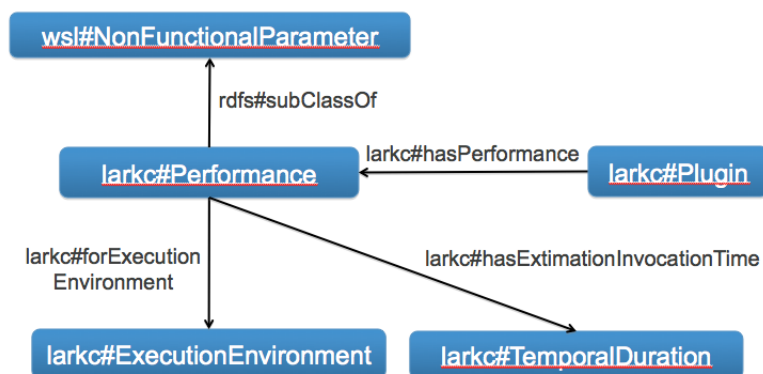


Figure 2.5: Performance

Listing 2.2 contains an example of how to describe performance aspects of a LarKC plug-in.

Listing 2.5: Performance example

```

@prefix larkc: <http://larkc.eu/plugin#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
    
```

```
@prefix sin: <http://www.sindice.com#> .

sin:SindiceIdentifier rdf:type larkc:Identifier ;
    larkc:hasPerformance sin:Performance .

sin:Performance rdf:type larkc:Performance ;
    larkc:hasEstimatedInvocationTime sin:Duration .

sin:Duration rdf:type larkc:TemporalDuration ;
    larkc:hasDurationValue 02500^^xsd:integer ;
    larkc:hasUnitOfDuration larkc:Milliseconds.
```

2.3 Scalability

Scalability in LarKC plug-in context refers to the scalability rate of the given plug-in. A graphical representation of the scalability fragment of the ontology is available in Figure 2.6.

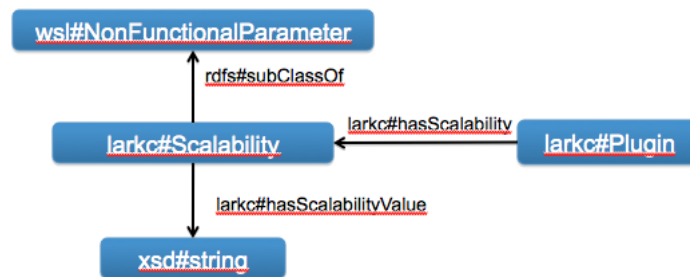


Figure 2.6: Scalability

Listing 2.3 contains an example of how to describe scalability aspects of a LarKC plug-in.

Listing 2.6: Scalability example

```
@prefix larkc: <http://larkc.eu/plugin#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix sti2: <http://www.sti2.at#> .

sti2:IrisReasoner rdf:type larkc:Reasoner ;
    larkc:hasScalability sti2:Scalability .

sti2:Scalability rdf:type larkc:Scalability ;
    larkc:hasScalabilityValue "linear"^^xsd:string .
```

2.4 Accuracy

The Accuracy part of the ontology includes all notions that are used to model the accuracy of the results produced by a particular plug-in, i.e. completeness, soundness, precision, recall, etc. The following subsections detail the accuracy concepts mentioned before.

2.4.1 Completeness and Soundness

A graphical representation of the Completeness and Soundness fragment of the ontology is available in Figure 2.7. This part of the ontology is used to specify whether completeness and soundness is required or achieved.

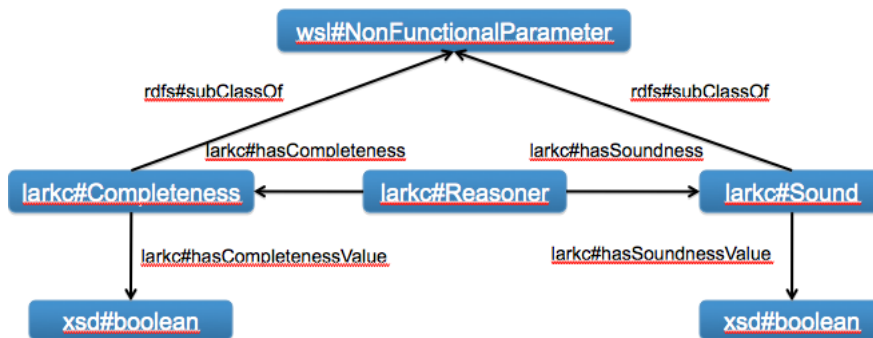


Figure 2.7: Completeness and Soundness

Listing 2.4.1 contains an example of how to describe Completeness and Soundness for a LarKC plug-in.

Listing 2.7: Completeness and Soundness example

```

@prefix larkc: <http://larkc.eu/plugin#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix sti2: <http://www.sti2.at#> .

sti2:IrisReasoner rdf:type larkc:Reasoner ;
    larkc:hasCompleteness sti2:Completeness ;
    larkc:hasSoundness sti2:Soundness .

sti2:Completeness rdf:type larkc:Completeness ;
    larkc:hasCompletenessValue "true"^^xsd:boolean .

sti2:Soundness rdf:type larkc:Soundness ;
    larkc:hasSoundnessValue "true"^^xsd:boolean .
    
```

2.4.2 Precision and Recall

A graphical representation of the Precision and Recall fragment of the ontology is available in Figure 2.8. This part of the ontology is used to specify whether completeness and soundness is required or achieved.

Listing 2.4.2 contains an example of how to describe Precision and Recall for a LarKC plug-in.

Listing 2.8: Precision and Recall example

```

@prefix larkc: <http://larkc.eu/plugin#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ex: <http://www.example.org#> .
    
```

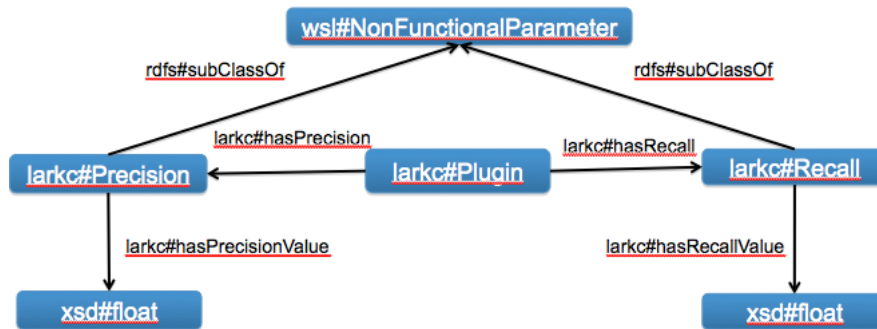


Figure 2.8: Precision and Recall

```

ex:SomeTransformer rdf:type larkc:Transformer ;
    larkc:hasRecall ex:Recall ;
    larkc:hasPrecision ex:Precision .

ex:Recall rdf:type larkc:Recall ;
    larkc:hasRecallValue "1.0"^^xsd:float .

ex:Precision rdf:type larkc:Precision ;
    larkc:hasPrecisionValue "1.0"^^xsd:float .
    
```

2.4.3 F-Measure, ROC Curve, AUC Curve

A graphical representation of the F-Measure, ROC Curve, AUC Curve fragment of the ontology is available in Figure 2.9. This part of the ontology is used to specify: the weighted harmonic mean of precision and recall i.e. F-MEasure, the ROC curve and the AUC Curve.

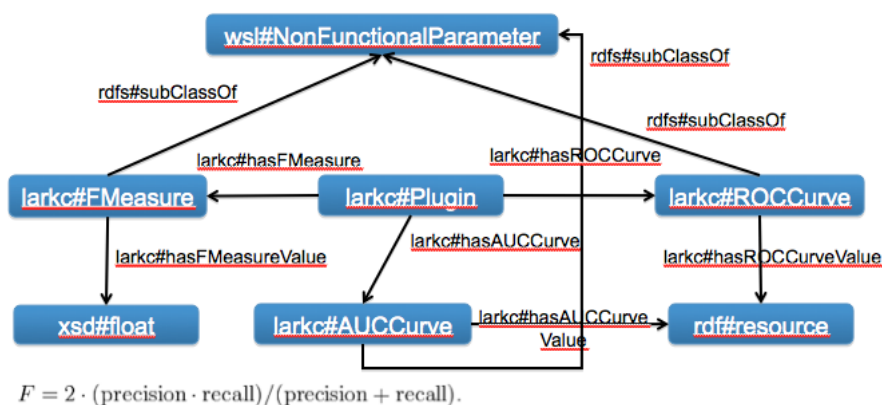


Figure 2.9: F-Measure, ROC Curve, AUC Curve

Listing 2.4.3 contains an example of how to describe F-Measure, ROC Curve, AUC Curve.

Listing 2.9: F-Measure, ROC Curve, AUC Curve example

```
@prefix larkc: <http://larkc.eu/plugin#> .
```

```

@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ex: <http://www.example.org#> .

ex:SomeTransformer rdf:type larkc:Transformer ;
    larkc:hasFMeasure ex:FMeasure ;
    larkc:hasROCCurve ex:ROCCurve ;
    larkc:hasAUCCurve ex:AUCCurve ;

ex:FMeasure rdf:type larkc:FMeasure ;
    larkc:hasFMeasureValue "1.0"^^xsd:float .

ex:ROCCurve rdf:type larkc:ROCCurve ;
    larkc:hasROCCurveValue ex:MyROCCurve .

ex:AUC rdf:type larkc:AUCCurve ;
    larkc:hasAUCValue ex:MyAUC .
    
```

2.5 Temporal

General temporal aspects of plug-ins can be described using the classes and properties contains in the temporal aspect of the ontology. A graphical representation of the temporal fragment of the ontology is available in Figure 2.10.

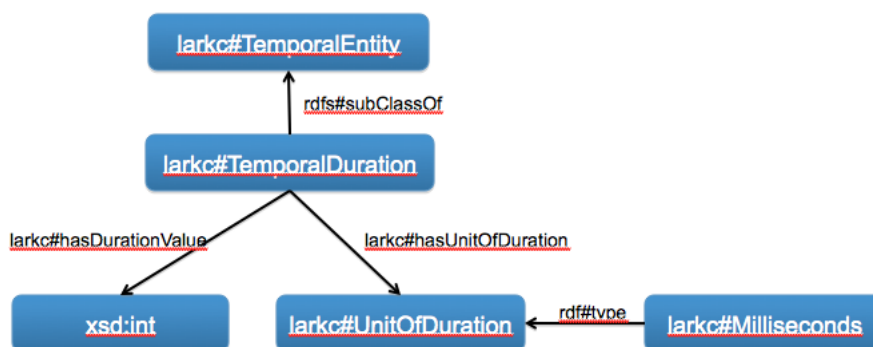


Figure 2.10: Temporal

The `larkc#TemporalDuration` class it is used for example when defining the estimated invocation time as part of the performance description of a plug-in.

2.6 Financial

The financial aspects that are modeled in the ontology involve terms related to cost-related and charging-related properties of a plug-in. A graphical representation of the financial fragment of the ontology is available in Figure 2.11.

Listing 2.6 contains an example of how to describe financial aspects of a LarKC plug-in.

Listing 2.10: Financial example

```

@prefix larkc: <http://larkc.eu/plugin#> .
    
```

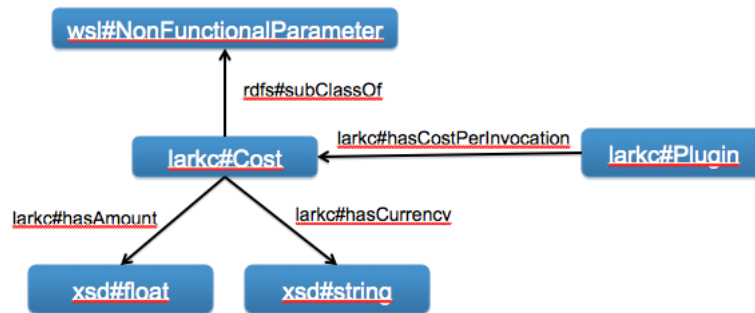


Figure 2.11: Financial

```

    @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
    @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
    @prefix ex: <http://www.example.org#> .

    ex:SomeTransformer rdf:type larkc:Transformer ;
        larkc:hasCostPerInvocation ex:Cost .

    ex:Cost rdf:type larkc:Cost ;
        larkc:hasAmount "0.10"^^xsd:float ;
        larkc:hasCurrency "EUR"^^xsd:string .
    
```

2.7 Utilities

2.7.1 Cardinality

A graphical representation of the cardinality fragment of the ontology is available in Figure 2.12. This part of the ontology is used to usually transformation cardinality (e.g. one to one, one to many, etc.).

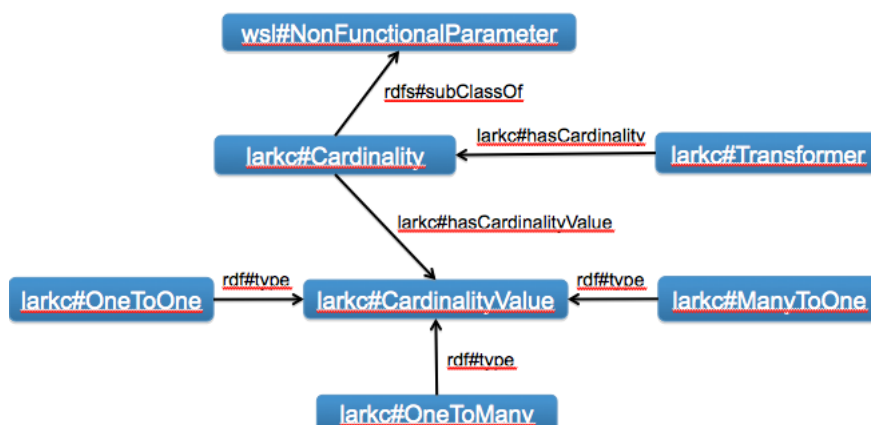


Figure 2.12: Cardinality

Listing 2.7.1 contains an example of how to describe cardinality.

Listing 2.11: Cardinality example



```
@prefix larkc: <http://larkc.eu/plugin#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix ex: <http://www.example.org#> .

ex:SomeTransformer rdf:type larkc:Transformer ;
    larkc:hasCardinality ex:Cardinality .

ex:Cardinality rdf:type larkc:Cardinality ;
    larkc:hasCardinalityValue larkc:OneToOne .
```

2.7.2 Limits

To describe minimum and maximum restrictions over the various requirements e.g. CPU, Memory, Deployment Factor we introduce a class called Limits which is depicted in Figure 2.13.

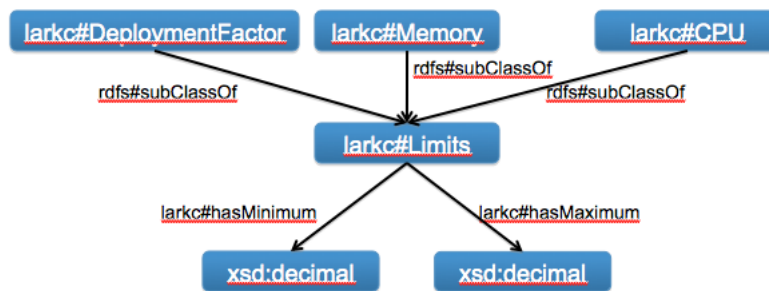


Figure 2.13: Limits

The full ontology is available in Listing 2.7.2.

Listing 2.12: LarKC Non-functional properties ontology

```

/*****
* LarKC Non-functional properties ontolgy
*****/
@prefix larkc: <http://lark.eu/plugin#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix wsl: <http://www.wsmo.org/ns/wsmo-lite#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .

/*****
* Infrastructure
*****/

larkc:ExecutionEnvironment rdfs:subClassOf wsl:
    NonFunctionalParameter .

larkc:Memory rdfs:subClassOf wsl:NonFunctionalParameter ;
    rdfs:subClassOf larkc:Limits .

larkc:hasMemory rdf:type rdf:Property ;
    rdfs:domain larkc:ExecutionEnvironment ;
    rdfs:range larkc:Memory .
```



```
larkc:hasMemoryValue rdf:type rdf:Property ;
                      rdfs:domain larkc:Memory ;
                      rdfs:range xsd:int .

larkc:hasMemoryUnit rdf:type rdf:Property ;
                     rdfs:domain larkc:Memory ;
                     rdfs:range larkc:MemoryUnit .

larkc:MemoryUnit rdf:type rdfs:Class .

larkc:Byte rdfs:subClassOf larkc:MemoryUnit .

larkc:CPU rdf:type rdfs:Class .

larkc:hasCPU rdf:type rdf:Property ;
              rdfs:domain larkc:ExecutionEnvironment ;
              rdfs:range larkc:CPU .

larkc:CPUSpeed rdf:type rdfs:Class .

larkc:hasCPUSpeed rdf:type rdf:Property ;
                  rdfs:domain larkc:CPU ;
                  rdfs:range larkc:CPUSpeed .

larkc:hasCPUSpeedValue rdf:type rdf:Property ;
                       rdfs:domain larkc:CPUSpeed ;
                       rdfs:range xsd:int .

larkc:hasCPUSpeedUnit rdf:type rdf:Property ;
                      rdfs:domain larkc:CPUSpeed ;
                      rdfs:range larkc:FrequencyUnit .

larkc:FrequencyUnit rdf:type rdfs:Class .

larkc:Hertz a larkc:FrequencyUnit .

larkc:requiresExecutionEnvironment rdf:type rdf:Property ;
                                   rdfs:domain larkc:Plugin ;
                                   rdfs:range larkc:
                                       ExecutionEnvironment .

larkc:requiresMemory rdf:type rdf:Property ;
                      rdfs:domain larkc:Plugin ;
                      rdfs:range larkc:Memory .

//DeploymentFactor
larkc:DeploymentFactor rdfs:subClassOf wsl:NonFunctionalParameter
;
                      rdfs:subClassOf larkc:Limits .

larkc:hasDeploymentFactor rdf:type rdf:Property ;
                          rdfs:domain larkc:Plugin ;
                          rdfs:range larkc:DeploymentFactor .

/*****
* Performance
*****/
```



```
//Performance
larkc:Performance rdfs:subClassOf wsl:NonFunctionalParameter .

larkc:hasPerformance rdf:type rdf:Property ;
                    rdfs:domain larkc:Plugin ;
                    rdfs:range larkc:Performance .

larkc:forExecutionEnvironment rdf:type rdf:Property ;
                             rdfs:domain larkc:Performance ;
                             rdfs:range larkc:
                                 ExecutionEnvironment .

larkc:hasEstimatedInvocationTime rdf:type rdf:Property ;
                                 rdfs:domain larkc:Performance ;
                                 rdfs:range larkc:
                                     TemporalDuration .

/*****
* Scalability
*****/
larkc:Scalability rdfs:subClassOf wsl:NonFunctionalParameter .

larkc:hasScalability rdf:type rdfs:Property ;
                    rdfs:domain larkc:Transformer ;
                    rdfs:range larkc:Scalability .

larkc:hasScalability rdf:type rdfs:Property ;
                    rdfs:domain larkc:Reasoner ;
                    rdfs:range larkc:Scalability .

larkc:hasScalabilityValue rdf:type rdf:Property ;
                          rdfs:domain larkc:Scalability ;
                          rdfs:range xsd:string .

/*****
* Accuracy
*****/

//Completeness
larkc:Completeness rdfs:subClassOf wsl:NonFunctionalParameter .

larkc:hasCompleteness rdf:type rdf:Property ;
                     rdfs:domain larkc:Reasoner ;
                     rdfs:range larkc:Completeness .

larkc:hasCompletenessValue rdf:type rdf:Property ;
                           rdfs:domain larkc:Completeness ;
                           rdfs:range xsd:boolean .

//Soundness
larkc:Soundness rdfs:subClassOf wsl:NonFunctionalParameter .

larkc:hasSoundness rdf:type rdf:Property ;
                  rdfs:domain larkc:Reasoner ;
                  rdfs:range larkc:Soundness .

larkc:hasSoundnessValue rdf:type rdf:Property ;
                        rdfs:domain larkc:Soundness ;
                        rdfs:range xsd:boolean .
```



```
//Recall
larkc:Recall rdfs:subClassOf wsl:NonFunctionalParameter .

larkc:hasRecall rdf:type rdf:Property ;
               rdfs:domain larkc:Plugin ;
               rdfs:range larkc:Recall .

larkc:hasRecallValue rdf:type rdf:Property ;
                    rdfs:domain larkc:Recall ;
                    rdfs:range xsd:float .

//Precision
larkc:Precision rdfs:subClassOf wsl:NonFunctionalParameter .

larkc:hasPrecision rdf:type rdf:Property ;
                  rdfs:domain larkc:Plugin ;
                  rdfs:range larkc:Precision .

larkc:hasPrecisionValue rdf:type rdf:Property ;
                       rdfs:domain larkc:Precision ;
                       rdfs:range xsd:float .

//F-Measure
larkc:FMeasure rdfs:subClassOf wsl:NonFunctionalParameter .

larkc:hasFMeasure rdf:type rdf:Property ;
                  rdfs:domain larkc:Plugin ;
                  rdfs:range larkc:FMeasure .

larkc:hasFMeasureValue rdf:type rdf:Property ;
                       rdfs:domain larkc:FMeasure ;
                       rdfs:range xsd:float .

//ROC curve
larkc:ROCCurve rdfs:subClassOf wsl:NonFunctionalParameter .

larkc:hasROCCurve rdf:type rdf:Property ;
                  rdfs:domain larkc:Plugin ;
                  rdfs:range larkc:ROCCurve .

larkc:hasROCCurveValue rdf:type rdf:Property ;
                       rdfs:domain larkc:ROCCurve ;
                       rdfs:range rdf:Resource .

//AUC - the area under ROC curve
larkc:AUC rdfs:subClassOf wsl:NonFunctionalParameter .

larkc:hasAUC rdf:type rdf:Property ;
             rdfs:domain larkc:Plugin ;
             rdfs:range larkc:AUC .

larkc:hasAUCValue rdf:type rdf:Property ;
                  rdfs:domain larkc:AUC ;
                  rdfs:range rdf:Resource .

/*****
* Temporal
*****/
larkc:TemporalEntity rdf:type rdfs:Class .
```



```
larkc:TemporalDuration rdf:subClassOf wsl:NonFunctionalParameter
;
                                rdfs:subClassOf larkc:TemporalEntity .

larkc:UnitOfDuration rdf:type rdfs:Class .

larkc:hasUnitOfDuration rdf:type rdf:Property ;
                        rdfs:domain larkc:TemporalDuration ;
                        rdfs:range larkc:UnitOfDuration .

larkc:hasDurationValue rdf:type rdf:Property ;
                       rdfs:domain larkc:TemporalDuration ;
                       rdfs:range xsd:float .

larkc:MinutesDuration a larkc:UnitOfDuration .

/*****
* Financial
*****/
larkc:Cost rdfs:subClassOf wsl:NonFunctionalParameter ;

larkc:hasAmount rdf:type rdf:Property ;
                rdfs:domain larkc:Cost ;
                rdfs:range xsd:float .

larkc:hasCurrency rdf:type rdf:Property ;
                  rdfs:domain larkc:Cost ;
                  rdfs:range xsd:string .

larkc:hasCostPerInvocation rdf:type rdfs:Property ;
                           rdfs:domain larkc:Plugin ;
                           rdfs:range larkc:Cost .

/*****
* Utilities
*****/

//Cardinality
larkc:Cardinality rdfs:subClassOf wsl:NonFunctionalParameter .

larkc:hasCardinality rdf:type rdfs:Property ;
                    rdfs:domain larkc:Transformer ;
                    rdfs:range larkc:Cardinality .

larkc:hasCardinalityValue rdf:type rdf:Property ;
                          rdfs:domain larkc:Cardinality ;
                          rdfs:range larkc:CardinalityValue .

larkc:CardinalityValue rdf:type rdfs:Class .

larkc:OneToOne a larkc:CardinalityValue .
larkc:OneToMany a larkc:CardinalityValue .
larkc:ManyToOne a larkc:CardinalityValue .

//Class Limits - something that has a min and a max value
larkc:Limits rdf:type rdfs:Class .

larkc:hasMinimum rdf:type rdf:Property ;
                 rdfs:domain larkc:Limits ;
```



```
        rdfs:range xsd:float .  
  
larkc:hasMaximum rdf:type rdf:Property ;  
                rdfs:domain larkc:Limits ;  
                rdfs:range xsd:float .
```



3. LarkKC datasets descriptions

This section contains the descriptions of two important LarkKC datasets, namely Linked Life Data¹ and FactForge². We used the void vocabulary to provide useful information about these LarkKC datasets. void³ is a vocabulary and a set of instructions that enables the discovery and usage of linked datasets. It is part of linking open data effort being primarily aimed at describing link datasets. Using void one can specify what the dataset is about, how it is linked with other data sets, how its data can be accessed (e.g. SPARQL endpoint or dump) and what vocabularies and ontologies are used in the dataset. Furthermore with void one can specify the number of statements, documents, the number of distinct resources, subjects, objects, etc.

The void description of the Linked Life Data and FactForge datasets in Turtle Syntax are available in Listing 3.1 and Listing 3.2.

Listing 3.1: void description of Linked Life Data dataset

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix void: <http://rdfs.org/ns/void#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix : <#> .

:LLD a void:Dataset ;
  foaf:homepage <http://linkedlifedata.com/> ;
  dcterms:title "Linked Life Data" ;
  dcterms:description "Linked Life Data (LLD) - Pathway and
    Interaction Knowledge Base" ;
  dcterms:subject <http://dbpedia.org/resource/Linked_Data> ;
  dcterms:subject <http://dbpedia.org/resource/Semantic_Web> ;
  dcterms:publisher <http://ontotext.com/> ;
  void:uriRegexPattern "http://linkedlifedata.com/.+" ;
  void:sparqlEndpoint <http://linkedlifedata.com/sparql> ;
  void:uriLookupEndpoint <http://linkedlifedata.com/> ;
  void:triples 6217184106 ; # 4486111031 explicit triples
  void:entities 853176424 ;

#Ontologies used
void:vocabulary <http://www.obofoundry.org/cgi-bin/detail.cgi?id=
  disease_ontology> ;
void:vocabulary <http://www.human-phenotype-ontology.org/index.php/
  hpo_home.html> ;
void:vocabulary <http://symptomontologywiki.igs.umaryland.edu/> ;

#Subsets of LinkedLifeData
void:subset :DrugBank ;
void:subset :Diseasome ;
void:subset :LinkedCT ;
void:subset :DailyMed ;
void:subset :SINDER ;
void:subset :BioGRID ;
void:subset :CellMap ;
void:subset :HPRD ;
```

¹<http://linkedlifedata.com/>

²<http://factforge.net/>

³<http://vocab.deri.ie/void/guide>



```
void:subset :HumanCYC ;
void:subset :IMID ;
void:subset :IntAct ;
void:subset :Reactome ;
void:subset :NCI-Nature ;
void:subset :MINT ;
void:subset :LHGDN ;
void:subset :Entrez-Gene ;
void:subset :PubMed ;
void:subset :UniProt ;
void:subset :ChEBI ;
void:subset :UMLS ;
void:subset :DiseaseOntology ;
void:subset :HumanPhenotypeOntology ;
void:subset :SymptomOntology .

:DrugBank
  a void:Dataset ;
  foaf:homepage <http://www.drugbank.ca/> ;
  rdfs:label "DrugBank" ;
  dcterms:title "A void Description of the DrugBank Dataset" ;
  dcterms:license <http://www.drugbank.ca/about> ;
  void:triples 517694 .

:Diseasome
  a void:Dataset ;
  foaf:homepage <http://diseasome.eu/> ;
  rdfs:label "Diseasome" ;
  dcterms:title "A void Description of the Diseasome Dataset" ;
  void:triples 144552.

:LinkedCT
  a void:Dataset ;
  foaf:homepage <http://data.linkedct.org/> ;
  rdfs:label "LinkedCT" ;
  dcterms:title "A void Description of the Linked Data Source
    of Clinical Trials (LinkedCT) Dataset" ;
  dcterms:license <http://www.clinicaltrials.gov/ct2/info/terms
    > ;
  void:triples 7110670 ;
  void:sparqlEndpoint <http://data.linkedct.org/sparql> .

:DailyMed
  a void:Dataset ;
  foaf:homepage <http://dailymed.nlm.nih.gov/dailymed/> ;
  rdfs:label "DailyMed" ;
  dcterms:title "A void Description of the DailyMed Dataset" ;
  dcterms:license <http://www.nlm.nih.gov/copyright.html> ;
  void:triples 192578 .

:SIDER
  a void:Dataset ;
  foaf:homepage <http://sideeffects.embl.de/> ;
  rdfs:label "SIDER" ;
  dcterms:title "A void Description of the Side Effect Resource
    (SIDER) Dataset" ;
  dcterms:license <http://creativecommons.org/licenses/by-nc-sa
    /3.0/> ;
  void:triples 106805 .
```



```
:BioGRID
  a void:Dataset ;
  foaf:homepage <http://thebiogrid.org/> ;
  rdfs:label "BioGRID" ;
  dcterms:title "A void Description of the BioGrid Dataset" ;
  void:triples 21017987 .

:CellMap
  a void:Dataset ;
  foaf:homepage <http://cancer.cellmap.org/cellmap/> ;
  rdfs:label "CellMap" ;
  dcterms:title "A void Description of the The Cancer Cell Map
    (CellMap) Dataset" ;
  void:triples 148350 .

:HPRD
  a void:Dataset ;
  foaf:homepage <http://www.hprd.org/> ;
  rdfs:label "HPRD" ;
  dcterms:title "A void Description of the Human Protein
    Reference Database" ;
  void:triples 1917111 .

:HumanCYC
  a void:Dataset ;
  foaf:homepage <http://humancyc.org/> ;
  rdfs:label "HumanCYC" ;
  dcterms:title "A void Description of the HumanCYC Dataset" ;
  dcterms:license <http://biocyc.org/all-reg.shtml> ;
  void:triples 300720 .

:IMID
  a void:Dataset ;
  foaf:homepage <http://www.sbcny.org/> ;
  rdfs:label "IMID" ;
  dcterms:title "A void Description of the IMID Dataset" ;
  dcterms:license <http://www.sbcny.org/data.htm> ;
  void:triples 81659 .

:IntAct
  a void:Dataset ;
  foaf:homepage <http://www.ebi.ac.uk/intact/> ;
  rdfs:label "IntAct" ;
  dcterms:title "A void Description of the IntAct Dataset" ;
  dcterms:license <http://creativecommons.org/licenses/by/2.5/>
    ;
  void:triples 16229745 .

:Reactome
  a void:Dataset ;
  foaf:homepage <http://reactome.org/> ;
  rdfs:label "Reactome" ;
  dcterms:title "A void Description of the Reactome Dataset" ;
  dcterms:license <http://fly.reactome.org/copyright.html> ;
  void:triples 698567 .

:NCI-Nature
  a void:Dataset ;
  foaf:homepage <http://pid.nci.nih.gov/> ;
  rdfs:label "NCI-Nature" ;
```



```
    dcterms:title "A void Description of the NCI-Nature Dataset"
    ;
    void:triples 614086 .

:MINT
a void:Dataset ;
foaf:homepage <http://mint.bio.uniroma2.it/mint/> ;
rdfs:label "MINT" ;
dcterms:title "A void Description of the Molecular
    INTeraction database (MINT) Dataset" ;
void:triples 20277714 .

:LHGDN
a void:Dataset ;
foaf:homepage <http://www.dbs.ifi.lmu.de/~bundschu/LHGDN.html
    > ;
rdfs:label "Literature-derived Human Gene-Disease Network (
    LHGDN)" ;
dcterms:title "A void Description of the Literature-derived
    Human Gene-Disease Network (LHGDN) Dataset" ;
dcterms:license <http://creativecommons.org/licenses/by-nc-sa
    /3.0/> ;
void:triples 316020 .

:Entrez-Gene
a void:Dataset ;
foaf:homepage <http://www.ncbi.nlm.nih.gov/sites/entrez?db=
    gene> ;
rdfs:label "Entrez-Gene" ;
dcterms:title "A void Description of the Entrez-Gene Dataset"
    ;
dcterms:license <http://www.nlm.nih.gov/copyright.html> ;
void:triples 140024275 .

:PubMed
a void:Dataset ;
foaf:homepage <http://www.ncbi.nlm.nih.gov/PubMed/> ;
rdfs:label "PubMedia" ;
dcterms:title "A void Description of the PubMed Dataset" ;
dcterms:license <http://www.nlm.nih.gov/copyright.html> ;
void:triples 1360059619 .

:UniProt
a void:Dataset ;
foaf:homepage <http://www.uniprot.org/> ;
rdfs:label "UniProt" ;
dcterms:title "A void Description of the UniProt Dataset" ;
dcterms:license <http://www.uniprot.org/help/license> ;
void:triples 1833277354 .

:ChEBI
a void:Dataset ;
foaf:homepage <http://www.ebi.ac.uk/chebi/> ;
rdfs:label "ChEBI" ;
dcterms:title "A void Description of the Chemical Entities of
    Biological Interest (ChEBI) Dataset" ;
dcterms:license <http://creativecommons.org/licenses/by/3.0/>
    ;
void:triples 322960 .
```



```
:UMLS
  a void:Dataset ;
  foaf:homepage <http://www.nlm.nih.gov/research/umls/> ;
  rdfs:label "UMLSMetathesaurus" ;
  dcterms:title "A void Description of the Unified Medical
    Language System (UMLS) Dataset" ;
  dcterms:license <http://wwwcf.nlm.nih.gov/umlslicense/snomed/
    license.cfm> ;
  void:triples 110568086 .

:DiseaseOntology
  a void:Dataset ;
  foaf:homepage <http://diseaseontology.sourceforge.net/> ;
  rdfs:label "DiseaseOntology" ;
  void:triples 144552 .

:HumanPhenotypeOntology
  a void:Dataset ;
  foaf:homepage <http://www.human-phenotype-ontology.org/index.
    php/hpo_home.html> ;
  rdfs:label "HumanPhenotypeOntology" ;
  void:triples 84378.

:SymptomOntology
  a void:Dataset ;
  foaf:homepage <http://symptomontologywiki.igs.umaryland.edu/>
    ;
  rdfs:label "SymptomOntology" ;
  void:triples 4163.
```

Listing 3.2: void description of FactForge dataset

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix dcterms: <http://purl.org/dc/terms/> .
@prefix void: <http://rdfs.org/ns/void#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix : <#> .

:FactForge a void:Dataset ;
  foaf:homepage <http://www.factforge.net/> ;
  dcterms:title "FactForge" ;
  dcterms:description "FactForge" ;
  dcterms:subject <http://dbpedia.org/resource/Linked_Data> ;
  dcterms:subject <http://dbpedia.org/resource/Semantic_Web> ;
  dcterms:publisher <http://ontotext.com/> ;
  void:uriRegexPattern "http://ldsr.ontotext.com/.+" ;
  void:sparqlEndpoint <http://ldsr.ontotext.com/sparql> ;
  void:uriLookupEndpoint <http://ldsr.ontotext.com/> ;
  void:triples 2237550383; # 880537138 inferred triples, 1357013225
    explicit triples
  void:entities 404796665;

#Ontologies used
void:vocabulary <http://purl.org/dc/> ;
void:vocabulary <http://www.w3.org/2004/02/skos/> ;
void:vocabulary <http://web.resource.org/rss/1.0/spec> ;
```



```

void:vocabulary <http://xmlns.com/foaf/0.1/> ;

#Subsets of LDSR
void:subset :DBpedia ;
void:subset :FreeBase ;
void:subset :GeoNames ;
void:subset :Umbel ;
void:subset :Wordnet ;
void:subset :FactBook ;
void:subset :Lingvoj ;
void:subset :Musicbrainz .

:DBpedia
  a void:Dataset ;
  foaf:homepage <http://dbpedia.org/> ;
  rdfs:label "DBpedia" ;
  dcterms:title "DBpedia" ;
  dcterms:license <http://www.gnu.org/licenses/fdl.html> ;
  void:triples 45464000 ;
  void:uriRegexPattern "http://dbpedia.org/.*" .

:FreeBase
  a void:Dataset ;
  foaf:homepage <http://rdf.freebase.com/> ;
  rdfs:label "FreeBase" ;
  dcterms:title "A void Description of the FreeBase Dataset" ;
  void:triples 504529000 ;
  void:uriRegexPattern "http://rdf.freebase.com/.*" .

:GeoNames
  a void:Dataset ;
  foaf:homepage <http://www.geonames.org/>;
  rdfs:label "GeoNames" ;
  dcterms:title "A void Description of the GeoNames Dataset" ;
  void:triples 216933000 ;
  void:uriRegexPattern "http://www.geonames.org/.*" .

:Umbel
  a void:Dataset ;
  foaf:homepage <http://www.umbel.org/>;
  rdfs:label "Umbel" ;
  dcterms:title "A void Description of the Umbel Dataset" ;
  void:triples 47374000 ;
  void:uriRegexPattern "http://www.umbel.org/.*" .

:Wordnet
  a void:Dataset ;
  foaf:homepage <http://wordnet.princeton.edu/> ;
  rdfs:label "Wordnet" ;
  dcterms:title "A void Description of the Worldnet Dataset" ;
  void:triples 11577000 ;
  void:uriRegexPattern "http://wordnet.princeton.edu/.*" .

:FactBook
  a void:Dataset ;
  foaf:homepage <http://www4.wiwiss.fu-berlin.de/factbook/>;
  rdfs:label "FactBook" ;
  dcterms:title "A void Description of the FactBook Dataset" ;
  void:triples 80000 ;

```



```
void:uriRegexPattern "http://www4.wiwiss.fu-berlin.de/
factbook/.*" .
```

```
:Lingvoj
```

```
a void:Dataset ;
foaf:homepage <http://www.lingvoj.org/>;
rdfs:label "Lingvoj" ;
dcterms:title "A void Description of the Lingvoj Dataset" ;
void:triples 883000 ;
void:uriRegexPattern "http://www.lingvoj.org/.*" .
```

```
:Musicbrainz
```

```
a void:Dataset ;
foaf:homepage <http://musicbrainz.org/>;
rdfs:label "Musicbrainz" ;
dcterms:title "A void Description of the Musicbrainz Dataset"
;
void:triples 466630000 ;
void:uriRegexPattern "http://musicbrainz.org/.*" .
```



4. Conclusion

This deliverable reported on the work that has been done in LarKC towards better language support for plug-in annotation. We described the updates and extensions of the initial LarKC plug-in annotation mechanism available in [1]. The focus was on the QoS aspects of plugins, the set of non-functional properties covered in the ontology being larger, much clear and better formalized than in the original version. Furthermore we illustrated how to create descriptions of datasets on two concrete LarKC datasets, namely Linked Life Data and FactForge. The voidD vocabulary was used to provide relevant information about these datasets such as the sets that are part of the dataset, number of triples, ontologies used in the dataset, etc.



REFERENCES

- [1] B. Bishop, D. Roman, J. Kopecky, G. Gallizo, and B. Fortuna, “Initial plug-in annotation language,” LarKC Project Deliverable, Tech. Rep. D1.3.1, 2009.
- [2] I. Toma, “Modeling and ranking semantic web services based on non-functional properties,” Ph.D. dissertation.