HAL
open science

Design of the second evaluation campaign

Stuart Wrigley, Raúl García Castro, Liliana Cabral, Cassia Trojahn dos Santos, Christian Meilicke, Lyndon Nixon, Mikalai Yatskevich

To cite this version:

Stuart Wrigley, Raúl García Castro, Liliana Cabral, Cassia Trojahn dos Santos, Christian Meilicke, et al.. Design of the second evaluation campaign. [Contract] 2011, pp.40. hal-00794905

HAL Id: hal-00794905

https://inria.hal.science/hal-00794905

Submitted on 26 Feb 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

hal-00794905, version 1 - 26 Feb 2013

SEALS

Semantic Evaluation at Large Scale

FP7 – 238975

D3.5 Design of the second evaluation campaign

Coordinator: Stuart N. Wrigley
With contributions from: Raúl García-Castro, Liliana Cabral, Cássia Trojahn dos Santos, Christian Meilicke, Lyndon Nixon, Mikalai Yatskevich
Quality Controller: Cássia Trojahn dos Santos
Quality Assurance Coordinator: Raúl García-Castro

Document Identifier:	SEALS/2011/D3.5/V1.0
Class Deliverable:	SEALS EU-IST-2009-238975
Version:	version 1.0
Date:	June 14, 2011
State:	final
Distribution:	public

Executive Summary

This deliverable summarises the work to date in the implementation of the SEALS methodology and design recommendations described in SEALS Deliverable 3.1 (García-Castro and Martín-Recuerda, 2009).

SEALS Deliverable D3.1 described the sequence of activities necessary to conduct the evaluation campaigns. This sequence was divided into four phases called *Initiation*, *Involvement*, *Preparation and Execution*, and *Dissemination*.

This deliverable covers the initial preparation of the second SEALS Evaluation Campaign. Chapters 2 and 3 describe the tasks performed during the *Initiation* and *Involvement* phases respectively.

Chapter 5 summarises the plan of work required in the remainder of the *Involvement* phase and then the *Preparation and execution* and *Dissemination* phases together with two timelines (one targeting participants and the other targeting the SEALS consortium members).

The deliverable also contains a number of appendices which contain two forms of announcements for the second SEALS Evaluation Campaign as well as the specific evaluation scenarios which each of the five technology areas will be addressing. Finally, the general SEALS Evaluation Campaign terms and conditions are provided.

hal-00794905, version 1 - 26 Feb 2013

Document Information

IST Project Number	FP7 – 238975	**Acronym**	SEALS
Full Title	Semantic Evaluation at Large Scale		
Project URL	`http://www.seals-project.eu/`		
Document URL			
EU Project Officer	Carmela Asero		

Deliverable	**Number**	3.5	**Title**	Design of the second evaluation campaign
Work Package	**Number**	3	**Title**	Evaluation Campaigns and Semantic Technology Roadmaps

Date of Delivery	**Contractual**	M25	**Actual**	30-6-11
Status	version 1.0		final ☒	
Nature	prototype ☐ report ☒ dissemination ☐			
Dissemination level	public ☒ consortium ☐			

Authors (Partner)	Stuart N. Wrigley (USFD), Raúl García-Castro (UPM), Liliana Cabral (OU), Cássia Trojahn dos Santos (INRIA), Christian Meilicke (UMA), Lyndon Nixon (STI2), Mikalai Yatskevich (OXF)			
Resp. Author	**Name**	Stuart N. Wrigley	**E-mail**	s.wrigley@dcs.shef.ac.uk
	Partner	University of Sheffield	**Phone**	+44 (114) 222 1880

Abstract (for dissemination)	This deliverable is concerned with the implementation of the second evaluation campaign based upon the methodology and design recommendations made in SEALS Deliverable D3.1 (García-Castro and Martín-Recuerda, 2009). This deliverable covers the initial preparation of the second SEALS Evaluation Campaign and describes the tasks that have been performed during the *Initiation* and *Involvement* phases. Furthermore, the deliverable describes the steps to be taken over the next few months and the actors who are responsible for those steps.
Keywords	evaluation campaign, methodology, design, guidelines, timeline

Version Log			
Issue Date	**Rev No.**	**Author**	**Change**
29-03-2011	1	Stuart Wrigley	Creation of report outline
23-05-2011	2	Stuart Wrigley	Campaign scenarios added
23-05-2011	3	Raúl García-Castro	Added scenarios and announcement for ontology engineering tools
2-06-2011	4	Stuart Wrigley	Final updates before QC
6-06-2011	5	Stuart Wrigley	QC updates
14-06-2011	6	Stuart Wrigley	QAC updates

hal-00794905, version 1 - 26 Feb 2013

Project Consortium Information

Participant's name	Partner	Contact
Universidad Politécnica de Madrid		Asunción Gómez-Pérez Email: asun@fi.upm.es
University of Sheffield		Fabio Ciravegna Email: fabio@dcs.shef.ac.uk
Forschungszentrum Informatik an der Universität Karlsruhe		Rudi Studer Email: studer@fzi.de
University of Innsbruck		Daniel Winkler Email: daniel.winkler@sti2.at
Institut National de Recherche en Informatique et en Automatique		Jérôme Euzenat Email: Jerome.Euzenat@inrialpes.fr
University of Mannheim		Heiner Stuckenschmidt Email: heiner@informatik.uni-mannheim.de
University of Zurich		Abraham Bernstein Email: bernstein@ifi.uzh.ch
Open University		Liliana Cabral Email: L.S.Cabral@open.ac.uk
Semantic Technology Institute International		Alexander Wahler Email: alexander.wahler@sti2.org
University of Oxford		Ian Horrocks Email: ian.horrocks@comlab.oxford.ac.uk

hal-00794905, version 1 - 26 Feb 2013

Table of Contents

List of figures 7

List of tables 8

1 Introduction 9

2 Tasks performed in the *Initiation* phase 11
2.1 Identify organizers 11
2.1.1 Membership 12
2.2 Organize evaluation scenarios 13

3 Tasks performed in the *Involvement* phase 14
3.1 Announce the evaluation campaign 14
3.2 Provide registration mechanisms 15

4 Community engagement 16
4.1 Incorporating external persons in SEALS 16
4.2 Promotion of SEALS and the SEALS Evaluation Campaign 17

5 Future activities and timeline 18
5.1 Tasks for the next period 18
5.1.1 *Initiation* phase 18
5.1.2 *Involvement* phase 18
5.1.3 *Preparation and Execution* phase 18
5.1.4 *Dissemination* phase 19
5.2 Timeline 20
5.2.1 Participants 20
5.2.2 SEALS Partners 20

References 21

A Evaluation scenarios 24
A.1 WP10: Ontology Engineering Tools 24
A.2 WP11: Storage and Reasoning Systems 24
A.3 WP12: Matching Tools 25
A.4 WP13: Semantic Search Tools 26
A.4.1 Automated Phase 26
A.4.2 User-in-the-loop Phase 26
A.5 WP14: Semantic Web Service Tools 26

B General announcement of the SEALS Evaluation Campaigns 27

hal-00794905, version 1 - 26 Feb 2013

C Long announcements of evaluation campaigns 29
C.1 Common to all tool areas . . . 29
C.1.1 Instructions on how to participate . . . 29
C.1.2 Evaluation Process . . . 29
C.1.3 How to find out more . . . 30
C.2 Ontology Engineering Tools . . . 30
C.3 Storage and Reasoning Systems . . . 31
C.4 Matching Tools . . . 32
C.5 Semantic Search Tools . . . 35
C.5.1 Introduction . . . 35
C.5.2 Criteria . . . 35
C.5.3 Two phase approach . . . 35
C.5.4 Take part . . . 36
C.6 Semantic Web Service Tools . . . 36
C.6.1 Evaluation Scenario Overview . . . 36
C.6.2 Evaluation datasets . . . 36
C.6.3 Take part . . . 37

D Evaluation campaign agreements 38
D.1 Terms of participation . . . 38
D.2 Use rights . . . 38
D.2.1 Rule applying to the non-public use of the data . . . 38
D.2.2 Rules applying to evaluation campaign participants . . . 38
D.2.3 Rules applying to people who did not participate in an evaluation campaign . . . 39
D.2.4 Rules applying to other cases . . . 39

hal-00794905, version 1 - 26 Feb 2013

List of Figures

1.1 The evaluation campaign process. 9

2.1 *Initiation* phase of the evaluation campaign process. 11

3.1 *Involvement* phase of the evaluation campaign process. 14

5.1 *Preparation and execution* phase of the evaluation campaign process. . 19
5.2 *Dissemination* phase of the evaluation campaign process. 20
5.3 Timeline of events for the second SEALS Evaluation Campaign. 22

hal-00794905, version 1 - 26 Feb 2013

List of Tables

2.1 Membership of the E.C.O.C.. 12
2.2 Membership of each E.C.E.C.. 13

hal-00794905, version 1 - 26 Feb 2013

1. Introduction

The SEALS project aims to create a lasting reference infrastructure for semantic technology evaluation (the SEALS Platform) and thus facilitate the continuous evaluation of semantic technologies at a large scale. The SEALS Platform will be an independent, open, scalable, extensible and sustainable infrastructure that will allow the evaluation of semantic technologies by providing an integrated set of evaluation services and test suites.

The SEALS project will take place in two 18-month stages and in each of these stages different evaluation campaigns will be performed for each of the technologies covered in the project. The SEALS Platform will be used in these evaluation campaigns and the results from the second SEALS Evaluation Campaign will be employed in creating semantic technology roadmaps that will identify sets of efficient and compatible tools for developing large-scale semantic applications.

This document focusses on the design of the second SEALS Evaluation Campaign and draws heavily upon the methodology and design recommendations made in SEALS Deliverable D3.1 (García-Castro and Martín-Recuerda, 2009). SEALS Deliverable D3.1 described the sequence of activities necessary to conduct the evaluation campaigns. This sequence, shown in Figure 1.1, is divided into four phases called *Initiation*, *Involvement*, *Preparation and Execution*, and *Dissemination*.

Figure 1.1: The evaluation campaign process.

The second SEALS Evaluation Campaign, which will address five core semantic technology areas, will run until April 2012 with final results being available by June 2012. This deliverable covers the initial preparation of the second SEALS Evaluation Campaign and starts describing the tasks performed during the *Initiation* and *Involvement* phases in chapters 2 and 3, respectively. Specifically, these chapters address the formation of two types of committee: the *Evaluation Campaign Organizing Committee* (E.C.O.C.) and a number of *Evaluation Campaign Executing Committees* (E.C.E.C.s) which will oversee the running of the second SEALS Evaluation Campaigns and specify the individual memberships of those committees. Furthermore, Chapter 3 describes the work undertaken to date by the E.C.E.C.s.

As can be seen from Figure 1.1, the *Preparation and Execution* phase runs concurrently with the *Involvement* phase. In Chapter 5 we summarise the plan of work required in the remainder of the *Involvement* phase, the *Preparation and execution* and *Dissemination* phases together with two timelines. Each timeline targets a different set of people associated with SEALS: the set of participants and the members of the SEALS consortium.

Chapter 4 describes a number of activities related to community engagement. This chapter addresses the role of external persons both in terms of their participation in the

hal-00794905, version 1 - 26 Feb 2013

E.C.O.C. and E.C.E.C.s as well as their potential participation in the wider research of the SEALS project. Chapter 4 also summarises efforts to promote SEALS and the second SEALS Evaluation Campaign.

The deliverable also contains a number of appendices which contain the specific evaluation scenarios which each of the five technology areas will be addressing. Finally, the general SEALS Evaluation Campaign terms and conditions are included in Appendix D.

hal-00794905, version 1 - 26 Feb 2013

2. Tasks performed in the *Initiation* phase

The Initiation phase comprises the set of tasks where the different people involved in the organization of the evaluation campaign and the evaluation scenarios are identified and where the different evaluation scenarios are defined. These tasks and their interdependencies are shown in Figure 2.1.

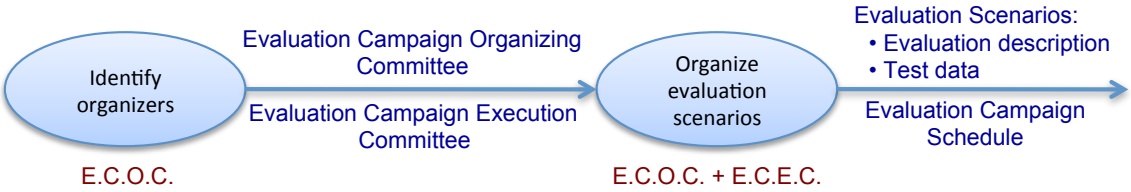

Figure 2.1: *Initiation* phase of the evaluation campaign process.

The first aspect of this task is the identification of the organisers; this is concerned with the creation of the two committees, described in Section 2.1, namely the E.C.O.C. and E.C.E.C.. The second aspect is the organisation of the evaluation scenarios, which is described in Section 2.2.

2.1 Identify organizers

The tasks of the evaluation campaign process are carried out by different actors according to the kind of roles that must be performed in each task. The SEALS Deliverable D3.1 (García-Castro and Martín-Recuerda, 2009) defined a number of different actors who will participate in the evaluation campaign process. This section presents the different kinds of actors involved in such process.

- **Evaluation Campaign Organizing Committee** (E.C.O.C.). The E.C.O.C. is in charge of the general organization and monitoring of the evaluation campaign. In SEALS there is one E.C.O.C. for all the evaluation campaigns and is composed of the SEALS Executive Project Management Board, the SEALS research work package leaders and other prominent external people.

- **Evaluation Campaign Executing Committee** (E.C.E.C.). The E.C.E.C. is in charge of organizing the evaluation scenarios that are performed in the evaluation campaign and of taking them to a successful end. In SEALS there will be at least one E.C.E.C. for each technology area (e.g., one E.C.E.C. for semantic search tool evaluation, one E.C.E.C. for matching tool evaluation, etc.).

- **Participants**. The evaluation campaign participants are tool providers or people with the permission of tool providers that participate with a tool in the evaluation campaign.

For the second evaluation campaign, effort has been made to expand the membership of the E.C.O.C. (and also the E.C.E.C.s) to include prominent, external members (see also Chapter 4). Therefore, membership of the E.C.O.C. and E.C.E.C. will be:

hal-00794905, version 1 - 26 Feb 2013

Table 2.1: Membership of the E.C.O.C..

Member	Affiliation	Project Role
Asunción Gómez-Pérez	UPM	EPMB Member
Fabio Ciravegna	USFD	EPMB Member
Jérôme Euzenat	INRIA	EPMB Member
Raúl García-Castro	UPM	Quality Assurance Coordinator
Lyndon Nixon	STI2	WP2 Leader
Catherina Burghart	FZI	WP10 Leader
Mikalai Yatskevich	OXF	WP11 Leader
Cássia Trojahn dos Santos	INRIA	WP12 Leader
Stuart Wrigley	USFD	WP13 Leader and WP3 Leader
Liliana Cabral	OU	WP14 Leader

- **Evaluation Campaign Organizing Committee** (E.C.O.C.). SEALS Executive Project Management Board (EPMB) and the SEALS research work package leaders plus any external personnel.
- **Evaluation Campaign Executing Committee** (E.C.E.C.). A subset of the relevant research work package as defined by the work package leader plus any external personnel.

2.1.1 Membership

E.C.O.C.

The E.C.O.C. is composed of the SEALS Executive Project Management Board, the SEALS research work package leaders and other prominent external people. The constituent members are shown in Table 2.1. Note that the E.C.O.C. also contains the WP3 Leader (coordinator of the evaluation campaign organisation work package) and the WP2 Leader (coordinator of the dissemination and community building work package). Furthermore, as specified in the Description of Work, Fabio Ciravegna (USFD) has the role of *Evaluation Campaigns Coordinator* and therefore will act as the chair of the E.C.O.C. with Stuart Wrigley (USFD; WP3 Leader) acting as his deputy.

At the time of writing this document, invitations to sit on the E.C.O.C. have also been extended to a number of external, well-respected members of the community. We are currently awaiting their responses.

- James A. Hendler, Tetherless World Senior Constellation Professor, Department of Computer Science and Cognitive Science Department, Rensselaer Polytechnic Institute (RPI), USA.
- Jeff Heflin, Associate Professor, Department of Computer Science and Engineering, Lehigh University, USA
- Frank Van Harmelen, Professor of Knowledge Representation and Reasoning, Department of Computer Science, VU University Amsterdam, Netherlands.

hal-00794905, version 1 - 26 Feb 2013

Table 2.2: Membership of each E.C.E.C..

WP	WP Name	Members	Affiliation
10	Ontology Engineering Tools	Raúl García-Castro Michael Schneider	UPM FZI
11	Storage and Reasoning Systems	Mikalai Yatskevich Michael Schneider Daniel Winkler	OXF FZI UIBK
12	Matching Tools	Jérôme Euzenat Christian Meilicke Ondřej Šváb-Zamazal	INRIA UMA Univ. of Economics, Prague, Czech Rep.
13	Semantic Search Tools	Stuart Wrigley Dorothee Reinhard	USFD UZH
14	Semantic Web Service Tools	Liliana Cabral Ning Li Daniel Winkler John Domingue	OU OU UIBK OU

- Enrico Motta, Professor of Knowledge Technologies, Knowledge Media Institute (KMi), The Open University, UK
- Mark Musen, Professor of Medicine (Biomedical Informatics), Stanford Center for Biomedical Informatics Research, Standford University, USA.
- Natasha Noy, Senior research scientist, Stanford Center for Biomedical Informatics Research, Standford University, USA.
- Matthias Klusch, Research Fellow, German Research Centre for Artificial Intelligence (DFKI), Germany.
- Birgitta König-Ries, Professor, Friedrich-Schiller-Universität Jena, Germany.

E.C.E.C.

Each E.C.E.C. contains a subset of the relevant WP personnel as shown in Table 2.2.

2.2 Organize evaluation scenarios

The E.C.E.C. is in charge of organizing the evaluation scenarios that are performed in the evaluation campaign and of taking them to a successful end. The evaluation scenarios identified by each E.C.E.C. are described in Appendix A.

hal-00794905, version 1 - 26 Feb 2013

3. Tasks performed in the *Involvement* phase

The *Involvement* phase comprises the set of tasks in which the evaluation campaign is announced and participants show their interest in participating by registering for the evaluation campaign. These tasks and their interdependencies are shown in Figure 3.1.

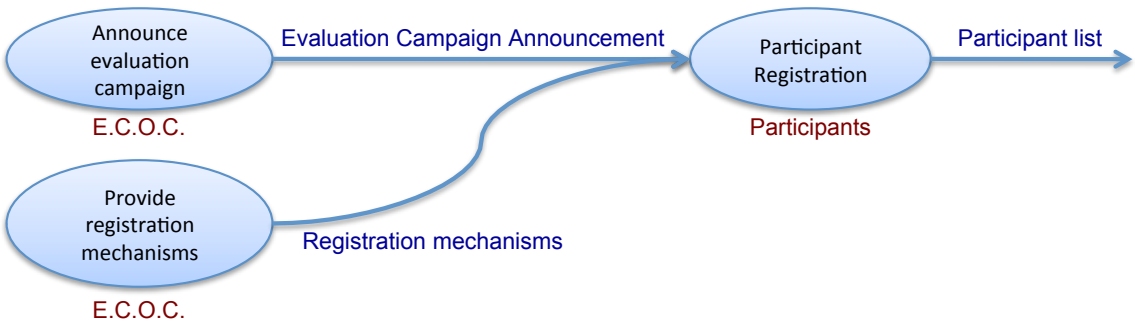

Figure 3.1: *Involvement* phase of the evaluation campaign process.

The tasks performed in this phase have been the preparation and dissemination of the evaluation campaign announcements in order to formally launch the second SEALS Evaluation Campaign at ESWC 2011 (first week of June). The mechanisms for participant registration, established for the first SEALS Evaluation Campaign, have been updated based upon feedback from each research work package.

3.1 Announce the evaluation campaign

As part of the *Involvement* phase, each E.C.E.C. has created two forms of public announcement: a 'short' and a 'long' announcement. These will be used both by WP2 and the E.C.E.C.s themselves to advertise the SEALS Evaluation Campaigns.

The short announcements have been combined to create a single announcement which gives a general description of the SEALS project and a broad introduction to each technology area's evaluation campaign (Appendix B). The target audience for this announcement is the semantic technology community in general and allied fields. Specifically, these distribution channels will be those maintained by STI International for such promotional activities as well as providers who have already registered an interest in SEALS via the SEALS Portal.

Each longer announcement will form the core information provided on the SEALS Portal for each technology area's second SEALS Evaluation Campaign. In addition, this information will also be emailed to researchers and tool developers known to be active in that particular tool field. These announcements give more details regarding the goals and evaluation scenarios involved in the evaluation campaign as well as information regarding how to participate. These announcements will be sent to the vendors identified as part of SEALS task T2.1 (described in D2.1 and enumerated on the SEALS private wiki) in which each of the work packages responsible for a particular technology area created a list of potential participants.

hal-00794905, version 1 - 26 Feb 2013

The announcements for each technology area's second SEALS Evaluation Campaign can be found in Appendix C with further details and up-to-date announcements being provided at http://www.seals-project.eu/seals-evaluation-campaigns/2nd-seals-evaluation-campaigns.

3.2 Provide registration mechanisms

Registration mechanisms were put in place on the SEALS Portal for the first SEALS Evaluation Campaign. These have been updated following consultation with each individual research workpackage (this consultation was conducted prior to the formal creation of each E.C.E.C.).

The general terms and conditions associated with participation in the SEALS Evaluation Campaigns (see appendix D) remain unchanged from the first SEALS Evaluation Campaign and are available on the SEALS Portal.

hal-00794905, version 1 - 26 Feb 2013

4. Community engagement

The SEALS Evaluation Campaigns are naturally community-oriented exercises with the overall goal of the project being to create a lasting infrastructure and best practice which persists after the funded duration of the project. At the core of the community involvement lies the participation of tool providers / developers who wish to evaluate their tools as well as tool adopters who will use the SEALS technology roadmaps and evaluation results to inform their technology decision processes.

However, it is also a goal of the SEALS project to involve technology providers and adopters in the design and execution of the SEALS Evaluation Campaigns themselves. This aspect is explicitly described in the methodology and design recommendations made in SEALS Deliverable D3.1 (García-Castro and Martín-Recuerda, 2009). In practice this relates to the involvement of persons external to the project in both the *Evaluation Campaign Organizing Committee* (E.C.O.C.) and each of the five technology areas' *Evaluation Campaign Executing Committees* (E.C.E.C.s). Such persons don't necessarily have to be drawn from technology developers or adopters, it is the intention of the SEALS consortium that high profile academic and industrial researchers also be involved. With respect to this, the SEALS Consortium has extended invitations to a number of external, well-respected members of the community (see Section 2.1). We are currently awaiting their responses.

Furthermore, it is hoped that existing evaluation efforts by members of the semantic community be aligned or even incorporated into the relevant SEALS Evaluation Campaigns. For instance, the SEALS Semantic Search Tool evaluation is currently working with the organisers of the Question Answering over Linked Data (QALD[1]) workshops with the intention of running their next campaign as part of SEALS. Additionally, Universidad Simón Bolívar is currently working with the SEALS Ontology Engineering Tool evaluation team with the intention of evaluating RDF storage systems.

Furthermore, the ability to incorporate external persons onto an E.C.E.C. provides a mechanism for this. In collaboration with the relevant E.C.E.C. and WP leader, the organiser can discuss the form of collaboration relevant to their evaluation effort. For instance, the SEALS Evaluation Campaign for Ontology Matching has an external member on its E.C.E.C. and it is likely that the SEALS Semantic Search Tool E.C.E.C. will also have an external member (representing QALD).

4.1 Incorporating external persons in SEALS

In addition to the inclusion of external representatives on the E.C.O.C. and the various E.C.E.C.s, the consortium has also put procedures in place for the creation of *Associated Partners*. Such persons could be identified by WP leaders on the basis that potential candidates can contribute concretely to the work done in SEALS, and specifically in their technology area. For instance, Universidad Simón Bolívar is currently working with WP11 as an Associated partner. The procedure that has been accepted by the SEALS consortium is the following:

[1] http://www.sc.cit-ec.uni-bielefeld.de/qald-1

hal-00794905, version 1 - 26 Feb 2013

1. WP leaders can either identify potential candidates which can contribute concretely to the work done in SEALS or be approached by such candidates.
2. Potential candidates formulate a short justification for their inclusion as associated partners including which SEALS work package they plan to contribute. This must be with the agreement of the respective WP leader.
3. The proposal is made by e-mail to the Project Management Board (PMB).
4. A decision will be made within one week based on the responses of the PMB members.
5. If accepted, a formal notification will be send to the new Associated Partner and their details, including planned contribution to SEALS, will be published on the SEALS Portal.

Associated partners can be invited to (parts of) SEALS meetings and participate in SEALS activities. Throughout their involvement with SEALS, the respective WP leader will act as the main liaison with the associated partner.

It must be noted that the granting of Associated Partner status does not infer any entitlement to SEALS funding nor any formal place in the SEALS consortium.

4.2 Promotion of SEALS and the SEALS Evaluation Campaign

SEALS Deliverable D2.1 (Nixon, 2009) describes the SEALS community building and dissemination plan which is split into two phases covering the lifespan of the project (M1-18, and M19-36). It identified three target groups to be addressed by the plan: the research community, the tool provider community and technology adopters. Clearly, the first phase is largely focused on the first two groups (research and tool provider communities), with the third group (technology adopters) receiving more focus in the second phase once the first SEALS Evaluation Campaign results are available. Since we now enter this second phase, and with the results of the first evaluation campaign available, SEALS has taken care to promote the campaign results by presentations (at ESTC 2010, and online, promoted from the SEALS webpage frontpage) and a whitepaper to be distributed to vendors. Updated promotional materials and a tutorial are used to introduce interested people to participation in the second campaign.

In particular, the ESWC 2011 conference, which took place on Crete from May 29 to June 2, 2011 was the launchpad for promoting the second evaluation campaign, and SEALS itself, to the semantic technology community. Following the SEALS tutorial on May 29, an ESWC booth in the reception area of the conference was available during all conference days and staffed by SEALS volunteers, giving attendees the opportunity to see and discuss more about SEALS and the evaluation campaign.

hal-00794905, version 1 - 26 Feb 2013

5. Future activities and timeline

This chapter describes the tasks — as identified in SEALS Deliverable D3.1 (García-Castro and Martín-Recuerda, 2009) — that will form the focus of work related to the organisation of the second SEALS Evaluation Campaign. This chapter has been split into two parts: the first part addresses the remaining phases and their associated tasks; the second part describes the timeline by which these tasks will be completed. Note that this latter part is further split into two to emphasise the differing responsibilities for SEALS Evaluation Campaign participants and SEALS Evaluation Campaign organisers (the E.C.E.C.s).

5.1 Tasks for the next period

This section summarises the tasks that need to be accomplished as part of the *Involvement*, *Preparation and execution* and *Dissemination* phases. These tasks will largely be the responsibility of the individual E.C.E.C.s for each of the five technology areas. However, some of the tasks described also require input from the technology providers / participants.

5.1.1 *Initiation* phase

The *Initiation* phase comprises the set of tasks where the different people involved in the organization of the evaluation campaign and the evaluation scenarios are identified and where the different evaluation scenarios are defined. This phase is complete except for the finalising of the E.C.O.C. with repect to the external members who will be invited to sit on the committee.

5.1.2 *Involvement* phase

The *Involvement* phase comprises the set of tasks in which the evaluation campaign is announced and participants show their interest in participating by registering for the evaluation campaign. These announcements have been created (see Sec. 3.1) and are available from the SEALS Portal. These, together with generic SEALS Evaluation Campaign publicity material were used to formally launch the SEALS Evaluation Campaign at ESWC 2011 (29 May – 2 June, 2011).

5.1.3 *Preparation and Execution* phase

The *Preparation and Execution* phase comprises the set of tasks that must be performed to insert the participating tools into the evaluation infrastructure, to execute each of the evaluation scenarios, and to analyse the evaluation results. These tasks and their interdependencies, shown in Figure 5.1, are the following:

1. **Provide evaluation materials**.
   The E.C.E.C. must provide to the registered participants all the evaluation materials needed in the evaluation, including:

hal-00794905, version 1 - 26 Feb 2013

(a) Instructions on how to participate.
(b) The evaluation description.
(c) The evaluation test data.
(d) The evaluation infrastructure.
(e) Any software needed for the evaluation.

2. **Insert tools**.
Once the participants have all the evaluation materials, they must insert their tools into the evaluation infrastructure and ensure that these tools are ready for the evaluation execution.

3. **Perform evaluation**.
The evaluation is executed over all the participating tools and the evaluation results of all the tools are collected.

4. **Analyse results**.
Once the evaluation results of all the tools are collected, they are analysed both individually for each tool and globally including all the tools. This results analysis must be reviewed in order to get agreed conclusions. Therefore, if the results are analysed by the E.C.E.C. then this analysis must be reviewed by the participants and vice versa, that is, if the results are analysed by the participants they must be reviewed by the E.C.E.C..

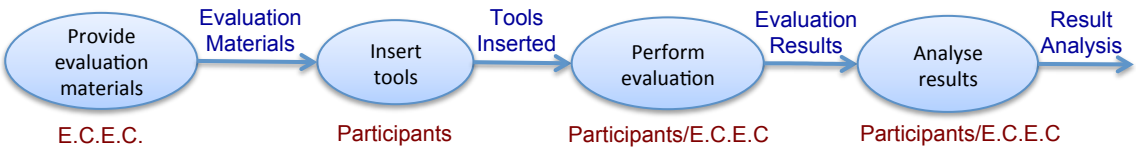

Figure 5.1: *Preparation and execution* phase of the evaluation campaign process.

5.1.4 *Dissemination* phase

The *Dissemination* phase comprises the set of tasks that must be performed to disseminate the evaluation campaign results by publicly presenting them and to make all the evaluation campaign results and resources available. The tasks that compose this phase can be performed either independently for each evaluation scenario or covering all the evaluation scenarios in each task. These tasks and their interdependencies, shown in Figure 5.2, are the following:

1. **Present results**.
The E.C.O.C., the E.C.E.C. and the participants will present and discuss the results of the evaluation campaign. It is envisaged that this will take the form of a workshop at ESWC 2012. The workshop will also be used to obtain feedback about the evaluation campaign.

hal-00794905, version 1 - 26 Feb 2013

2. **Publish results**.
   The E.C.O.C., the E.C.E.C. and the participants will publish the results of the evaluation campaign and of each of the tools either as workshop proceedings or journal special issues.

3. **Finalize**.
   All the evaluation resources used in the evaluation campaign will be made public. The final report — SEALS Deliverable D3.6 — will include the results of the campaign as well as feedback for improving future campaigns beyond the funded period of the SEALS project. In addition to this, SEALS Deliverable D3.7 will identify the semantic technology roadmaps showing inter-tool compatibilities. Finally, SEALS Deliverable D3.8 will provide an updated version of the benchmarking methodology in light of the lessons learned from both the first and second evaluation campaigns.

Figure 5.2: *Dissemination* phase of the evaluation campaign process.

5.2 Timeline

This section provides two timelines describing the work planned until the completion of the second SEALS Evaluation Campaign. Each timeline targets a different set of people associated with SEALS: the set of participants and the members of the SEALS consortium. However, each timeline should not be read in isolation: they are inherently interdependent and only together do they give a full description of the project's future work. Both timelines are shown Figure 5.3 with more details in Sections 5.2.1 and 5.2.2.

5.2.1 Participants

- June 2011 onwards: Register for participation at the SEALS Portal
- August 2011: Obtain evaluation materials
- October 2011 onwards: Insert tools into the SEALS Platform
- November 2011 onwards: In conjunction with E.C.E.C., perform evaluation
- March 2012: Result analysis
- Early June 2012: Workshop at ESWC 2012

5.2.2 SEALS Partners

- June 2011: Evaluation Campaign Announced.

hal-00794905, version 1 - 26 Feb 2013

  - The E.C.O.C., led by the WP2 leader, will disseminate the general announcement including all evaluation campaigns to the existing distribution channels (mailing lists, blogs, SEALS Portal, etc.).
  - Each E.C.E.C. will send technology area-specific announcements to all the vendors identified for T2.1 and T2.5 announcing details of their evaluation campaigns including timelines aimed at the potential participants.

- July 2011: Each E.C.E.C. makes evaluation materials available.
- August 2011: Each E.C.E.C. publishes final release (FR) version of v1.0 evaluation software.
- September 2011: Each E.C.E.C. inserts final test data into the SEALS Platform.
- September 2011: Tools may be inserted.
- End October 2011: Each E.C.E.C. inserts the final evaluation scenarios (description + workflow) into the SEALS Platform. This will include the provision of any software needed for the evaluation, inserted/connected into/with the SEALS Platform.
- November 2011: Evaluations conducted by each E.C.E.C. (possibly directly including participants, e.g., WP13).
- January 2012: Each E.C.E.C. publishes beta version of v2.0 evaluation software.
- February 2012: Each E.C.E.C. inserts final results analysis tools into the SEALS Platform.
- March 2012: Evaluation results ready.
- March 2012: Result analysis begins.
- May 2012: Result analysis complete.
- Early June 2012: Workshop at ESWC 2012.
- June 2012: Each E.C.E.C. makes evaluation resources public.
- June 2012: Each E.C.E.C. publishes results of evaluation.
- June 2012: E.C.O.C. and E.C.E.C.s produce final report D3.6.
- June 2012: Roadmap report (D3.7) and updated methodology (D3.8).
- June 2012: Each E.C.E.C. publishes final release (FR) version of v2.0 evaluation software.

hal-00794905, version 1 - 26 Feb 2013

hal-00794905, version 1 - 26 Feb 2013

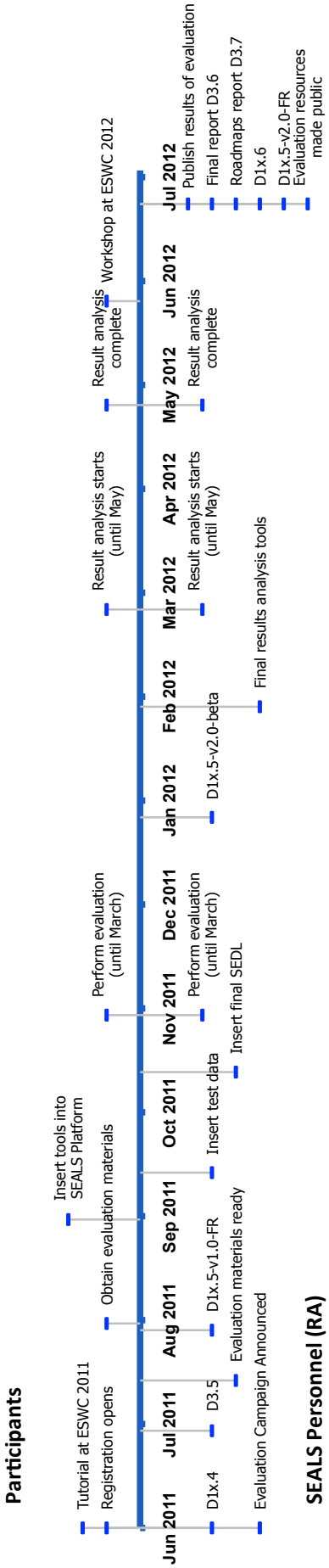

Figure 5.3: Timeline of events for the second SEALS Evaluation Campaign.

References

R. García-Castro and F. Martín-Recuerda. D3.1 SEALS Methodology for Evaluation Campaigns v1. Technical report, SEALS Consortium, 2009.

L. Nixon. D2.1 Plans for community building and dissemination v1. Technical report, SEALS Consortium, 2009.

hal-00794905, version 1 - 26 Feb 2013

A. Evaluation scenarios

A.1 WP10: Ontology Engineering Tools

The following naming scheme has been adopted to enable easy identification of evaluation campaign scenarios:

```
<ToolType> <Name/Criterion> <Year>
```

The evaluation scenarios are as follows (OET = ontology engineering tool):

- OET Conformance 2011
  Testing conformance of OETs according to a language specification.
  - RDF(S)
  - OWL Lite
  - OWL DL
  - OWL Full
  - OWL 2 Full
  - Ontology Design Pattern
- OET Interoperability 2011
  Testing interoperability between OETs when using an interchange language.
  - RDF(S)
  - OWL Lite
  - OWL DL
  - OWL Full
  - OWL 2 Full
  - Ontology Design Pattern
- OET Scalability 2011
  Testing loading times for handling large ontologies with OETs.
  - Real-world
  - LUBM

A.2 WP11: Storage and Reasoning Systems

The following naming scheme has been adopted to enable easy identification of evaluation campaign scenarios:

```
<ToolType> <Name/Criterion> <Year>
```


hal-00794905, version 1 - 26 Feb 2013

The evaluation scenarios are as follows (DLBS: Description logic based system; RDF SS: RDF storage system; RDF RS: RDF reasoning system):

- DLBS Classification 2011
- DLBS Class satisfiability 2011
- DLBS Ontology satisfiability 2011
- DLBS Logical entailment 2011
- DLBS Instance retrieval 2011
- RDF RS Specification-Level Evaluation 2011
- RDF RS Pragmatic-Feature-Level Evaluation 2011
- RDF SS Dataset load 2011
- RDF SS Query evaluation 2011

A.3 WP12: Matching Tools

In 2010 we used a web-based evaluation service, which did not allow to measure runtime efficiency. In contrast to the 2010 evaluation all scenarios will be executed on the SEALS platform in 2011. The following naming scheme has been adopted to enable easy identification of evaluation campaign scenarios:

```
<ToolType> <Name/Criterion> <Year>
```

The evaluation scenarios are as follows (MT = Matching Tool):

- MT Benchmark 2011
  Criteria: conformance with expected results, efficiency in terms of execution time, and interoperability.

- MT Anatomy 2011 (refined version of MT Anatomy 2010)
  Criteria: conformance with expected results, efficiency in terms of execution time, and interoperability.

- MT Conference 2011
  Criteria: conformance with expected results, efficiency in terms of execution time, interoperability, and alignment coherence.

Currently other datasets are analyzed and might be used to complement these scenarios. For example, we have developed a new test data generator, which will probably be used to create an additional Benchmark dataset that is build on a different reference ontology compared to MT Benchmark 2011. Moreover, we have contacted the organizers of the OAEI directory and instance matching track. The corresponding datasets might be used in additional SEALS scenarios. A final decision about the usage of these datasets will be made soon.

hal-00794905, version 1 - 26 Feb 2013

A.4 WP13: Semantic Search Tools

The following naming scheme has been adopted to enable easy identification of evaluation campaign scenarios:

`<ToolType> <Phase> <Name/Criterion> <Year>`

The evaluation scenarios are listed below (SST = Semantic Search Tool). It should be noted that participants of the Semantic Search Evaluation Campaign do *not* need to participate in both phases.

A.4.1 Automated Phase

1. SST Automated Search Performance 2011
   The tool's core search quality in terms of precision, recall, etc. The tool's scalability will be considered according to measures such as processing / load time.

A.4.2 User-in-the-loop Phase

1. SST User Usability 2011
   - How do the end-users react to the tool's query language? Do they like the tool? Are they able to express their questions effectively and fast? Is the language easy to understand and learn? These aspects of user satisfaction will be assessed in the questionnaires.
   - Interpretation of the correlations between the user demographics and the measures.
2. SST User Query expressiveness 2011
   - Testing the formal (by asking the test subjects in the questionnaire), as well as, the practical expressiveness (by running queries in order to test the actual coverage and robustness) of the tool's query language, i.e. if the tool is able to answer complex queries.
   - Were all queries answered by the tool; i.e., could an answer be found?

A.5 WP14: Semantic Web Service Tools

The following naming scheme has been adopted to enable easy identification of evaluation campaign scenarios:

`<ToolType> <Name/Criterion> <Year>`

The evaluation scenarios are as follows (SWS = Semantic Web Services):

- SWS Tool Discovery Evaluation 2011
  Retrieval and execution time performance evaluation of SWS tools for the discovery of Web Services based on their semantic annotations.

hal-00794905, version 1 - 26 Feb 2013

B. General announcement of the SEALS Evaluation Campaigns

In order to announce the various SEALS Evaluation Campaigns, the E.C.O.C. will distribute a general communication using any mechanism available (e.g., mailing lists, blogs, etc.) with the goal of reaching the developers of the tools that are targeted by the evaluation scenarios. This appendix contains this announcement and incorporates short explanations of each area's campaign and targets the general semantic web community.

Dear Sir/Madam,

We are pleased to announce the second world-wide SEALS Evaluation Campaign for semantic technologies which will take place during Autumn 2011 / Spring 2012. The SEALS project has created a lasting reference infrastructure for automated semantic technology evaluation (the SEALS Platform) and thus facilitating the continuous evaluation of semantic technologies at a large scale with minimal effort from participants.

We cordially invite you to participate in the SEALS Evaluation Campaign in one or more of the five core areas shown below. Participation is open to anyone who is interesting in benchmarking a semantic technology tool. Detailed information regarding each area's campaign together with terms and conditions and general information about SEALS can be found on the SEALS Portal at http://www.seals-project.eu.

SEALS evaluations will address five core semantic technology areas:

Ontology Engineering Tools: addresses the ontology management capabilities of semantic technologies in terms of their ontology language conformance, interoperability and scalability. The main tools targeted are ontology engineering tools and ontology management frameworks and APIs; nevertheless, the evaluation is open to any other type of semantic technology.

Storage and Reasoning Systems: assesses a reasoner's performance in various scenarios resembling real-world applications. In particular, their effectiveness (comparison with pre-established 'golden standards'), interoperability (compliance with standards) and scalability are evaluated with ontologies of varying size and complexity.

Matching Tools: builds on previous matching evaluation initiatives (OAEI campaigns) and integrates the following evaluation criteria: (a) conformance with expected results (precision, recall and generalizations); (b) performance in terms of memory consumption and execution time; (c) interoperability, measuring the conformance with standard such as RDF/OWL; and (d) measuring the coherence of the generated alignments.

Semantic Search Tools: evaluated according to a number of different criteria including query expressiveness (means by which queries are formulated within the tool) and scalability. Given the interactive nature of semantic search tools, a core interest in this evaluation is the usability of a particular tool (effectiveness, efficiency, satisfac-

hal-00794905, version 1 - 26 Feb 2013

tion).

Semantic Web Service Tools: focuses on activities such as discovery, ranking and selection. In the context of SEALS, we view a SWS tool as a collection of components (platform services) of the Semantic Execution Environment Reference Architecture (SEE-RA). Therefore, we require that SWS tools implement one or more SEE APIs in order to be evaluated.

hal-00794905, version 1 - 26 Feb 2013

C. Long announcements of evaluation campaigns

Once the different evaluation scenarios are defined, the E.C.O.C. must announce the evaluation campaign using any mechanism available (e.g., mailing lists, blogs, etc.) with the goal of reaching the developers of the tools that are targeted by the evaluation scenarios. This appendix contains the full announcements of each area's evaluation campaign. These will target tool providers and researchers in specific areas.

C.1 Common to all tool areas

C.1.1 Instructions on how to participate

Tool developers are cordially invited to participate in the SEALS Evaluation Campaign in one or more of the five core areas. Participation is open to developers interested in evaluating their tool or to anyone who wants to evaluate a certain tool.

From the tool provider's perspective, the SEALS Evaluation Campaign has been designed to require minimal effort on their part. For the majority of SEALS Evaluation Campaigns, the evaluation process is fully automated from the point at which the tool has been uploaded to the SEALS Platform. The SEALS Platform will manage all aspects of the evaluation. Once the evaluation has been completed, the results and analyses will be available from the SEALS Portal. In order to facilitate interaction between your tool and the SEALS Platform, the SEALS teams from each technology area have defined an easy to use API. Details about each technology area's API is available from their SEALS Evaluation Campaign's Portal page (http://www.seals-project.eu).

C.1.2 Evaluation Process

The timeline described in the following holds for all technology areas.

1. Preparation Phase (June 2011 – Nov 2011)
   The first stage of this phase is the release of the detailed SEALS Evaluation Campaign materials. These will consist of

   (a) Instructions on how to participate.
   (b) The evaluation description.
   (c) The evaluation test data.
   (d) Any software needed for the evaluation.

   Once these materials have been made available, participants will be able to submit and install their tools on the SEALS Platform. During this phase, the tool provider assisted by the E.C.E.C. have to verify that the tools submitted for evaluation run correctly on the SEALS Platform. In addition, first tests can be run with datasets associated with that SEALS Evaluation Campaign.

2. Execution Phase (Nov 2011 – March 2012)
   Participants can submit different versions of their tool and use the SEALS Plat-

hal-00794905, version 1 - 26 Feb 2013

form to evaluate them at any moment. When they feel confident that it is correctly configured, they can submit the final version to the system. It will then be evaluated and the raw results of the evaluation experiments will be stored in the SEALS Results Repository.

3. Analysis Phase (March – June 2012)
   The evaluation results will be automatically generated by the SEALS Platform using the evaluation criteria identified by your tool's technology area E.C.E.C.. This process will generate different data representations, such as graphs and tables for visualizing recall/precision, time execution, and comparisons between the different campaign participants. The results will be publicly accessible in the SEALS Portal after the evaluation phase has been conducted.

C.1.3 How to find out more

Detailed information regarding the SEALS Evaluation Campaign together with terms and conditions and general information about SEALS can be found on the SEALS Portal at http://www.seals-project.eu.

Each technology area's SEALS Evaluation Campaign is organised and executed by a SEALS *Evaluation Campaign Executing Committee* (E.C.E.C.). For more information on the SEALS Evaluation Campaign for this specific technology area, please contact the E.C.E.C.:
INSERT YOUR E.C.E.C. TABLE HERE

C.2 Ontology Engineering Tools

When dealing with the many existing semantic technologies, some questions arise sooner or later:

- Tool A is able of managing OWL DL ontologies but, up to what extent can it manage OWL Full ontologies?
- I am using an OWL Full ontology in Tool B and I want to use it in Tool C, which only supports OWL Lite. Can I make it with a minimal loss of information?
- Someone recommended me Tool D, but I need to manage very big ontologies. Can this tool make it efficiently? If not, which one can?

The SEALS Yardsticks For Ontology Management is an evaluation campaign that comprises a set of evaluations defined with the goal of evaluating the ontology management capabilities of semantic technologies in order to answer those questions.

The main tools targeted for these evaluations are ontology engineering tools and ontology management frameworks and APIs; nevertheless, the evaluation is open to any other type of semantic technology.

In the previous evaluation campaign (http://www.seals-project.eu/seals-evaluation-campa 1st-evaluation-campaigns/ontology-engineering-tools) we evaluated ontology development tools in terms of their conformance and interoperability (using RDF(S), OWL Lite, OWL DL, and OWL Full) and their scalability (using both real-world ontologies and synthetically-generated ones).

hal-00794905, version 1 - 26 Feb 2013

In this edition of the evaluation campaign we will also evaluate conformance, interoperability and scalability, using the same test data as in the previous edition and extending it with

- Test data based in ontology design patterns and OWL 2 Full test data for conformance and interoperability evaluations.
- New real-world and synthetic test data for scalability evaluations.

The evaluation campaign will take place by the end of 2011.

Participation is open to developers interested in evaluating their tool or to anyone who wants to evaluate a certain tool.

Participants are just expected to collaborate in the connection of their tool with the SEALS Platform, which will be the infrastructure that will run all the evaluations automatically. Besides checking their results and comparing with others, once the tool is connected to the SEALS Platform participants will also be able to run the evaluations on their own with these and future test data.

If you want to participate, register your tool in the SEALS Portal (http://www.seals-project.eu) and stay tuned to the evaluation campaign web page where you can find detailed descriptions of the evaluations that we will perform and the latest information and results of the evaluation campaign.

If you have any question or comment about the evaluation campaign, please contact us.

- Raúl García-Castro (*rgarcia@fi.upm.es*).
- Michael Schneider (*schneid@fzi.de*).

We count on your participation!

C.3 Storage and Reasoning Systems

Semantic technologies are at the heart of the future Web providing ways to express knowledge and data so that it can be properly exploited. These technologies will empower a new class of Information and Communication Technologies much more scalable, interoperable, and with a higher degree of process automation support that will fulfil the needs of an emergence market that will exceed $10 billion by 2010.

Description logic based systems (DLBSs) evaluation aims at assessing interoperability, performance and scalability of one of the core building blocks of semantic technologies vision namely description logic reasoners. Resource description framework reasoning systems (RDF RS) evaluation is targeted at assessing interoperability and performance of RDF reasoners. Resource description framework storage systems (RDF SS) evaluation assesses efficiency and effectiveness of RDF storage systems.

This evaluation, thus, serves several purposes:

- Helps researches and users to select appropriate technologies;
- Gives an objective comparison among existing technologies;
- Stimulates technologies advancement through continuous evaluation.

hal-00794905, version 1 - 26 Feb 2013

hal-00794905, version 1 - 26 Feb 2013

The overall objective of evaluation campaign is to evaluate DLBSs standard inference services: classification, class satisfiability, ontology satisfiability, logical entailment and instance retrieval. The challenge uses a set of set of state of the art ontologies for evaluation. The set includes OWL 2 test cases repository, ontologies from Gardiner suite, various versions of the GALEN ontology, ontologies that have been created in EU funded projects SEMINTEC, VICODI, AEO and SNOMED CT ontology. DLBSs are expected to support OWL 2 language and provide interface to their functionalities through a set of evaluation interfaces defined in the SEALS Platform (http://www.seals-project.eu/seals-platform).

The RDF RS evaluation includes inconsistency checking and entailment checking inference services. Test data includes OWL 2 Conformance suite with over 1000 of specification-level and feature-level tests and artificially generated RDF datasets.

The RDF SS evaluation includes dataset loading and query assessment inference services. The real-world datasets include DBpedia, YAGO, linkedCT, US congress votes and set of SPARQL queries for each of these datasets. Artificially generated datasets include LUBM and Berlin SPARQL benchmark. RDF SS are expected to support SPARQL query language.

The evaluation is administered by the EU FP7-238975 SEALS – Semantic Evaluation at Large Scale project http://www.seals-project.eu/.

If you want to participate, register your tool in the SEALS Portal (http://www.seals-project.eu) and stay tuned to the evaluation campaign web page where you can find detailed descriptions of the evaluations that we will perform and the latest information and results of the evaluation campaign.

If you have any question or comment about the evaluation campaign, please contact us.

- Mikalai Yatskevich (*mikalai.yatskevich@comlab.ox.ac.uk*).
- Michael Schneider (*schneid@fzi.de*).
- Daniel Winkler (*daniel.winkler@sti2.at*).

We count on your participation!

C.4 Matching Tools

The SEALS evaluation campaign, built on the previous matching evaluation initiatives (Ontology Alignment Evaluation Initiative - OAEI), aims at evaluating the competence of matching systems with respect to different evaluation criteria. The evaluation will focus on demonstrating the feasibility and benefits of automating matching evaluation.

Important modifications for 2011

SEALS is supporting the OAEI organizers in running several OAEI 2011 tracks on the SEALS platform. This year, participants are required to upload their tools into the SEALS platform instead of implementing web service interfaces. Participants must provide their tools in a ZIP file format. A tutorial is provided (details can be found below) in order to explain how to prepare, upload, and package the tools in the required

format. The webbased evaluation service used in the 2010 campaign, now integrated into the SEALS portal (https://www.seals-project.eu/ontology-matching-evaluation-ui), is still available for testing purposes but it will be not be used in the campaign to generate the official results.

Overview of the Evaluation Campaign

In 2011 we focus on the following criteria:

- Conformance against a reference alignment (precision, recall, f-measure);
- Interoperability, measuring the conformance with standard such as RDF/OWL;
- Coherence of alignments generated by matching systems;
- Efficiency in terms of runtime (elapsed runtime)

Notice that efficiency metrics are for the first time part of the evaluation compared to previous OAEI evaluations. Measuring efficiency was already planned for 2010, however, we finally had to postpone this issue to 2011. Running tools on the platform in 2011 allows us now to include runtime efficiency criteria (and also to take care of the reproducibility of results in a controlled environment).

Evaluation scenarios and datasets

The evaluation campaign contains at least three different scenarios, where the tools will be evaluated according to common datasets and criteria.

Scenario 1: Benchmark The goal of this benchmark series is to identify the areas in which each matching algorithm is strong or weak. The test is based on one particular ontology dedicated to the very narrow domain of bibliography and a number of alternative ontologies on the same domain for which alignments are provided.

- Criteria: conformance with expected results.
- URL in SEALS testrepository http://seals.sti2.at/tdrs-web/testdata/persistent/Benchmark+Testsuite/2010full/suite

Scenario 2: Anatomy The anatomy real world case is about matching the Adult Mouse Anatomy (2744 classes) and the NCI Thesaurus (3304 classes) describing the human anatomy. Both ontologies are relatively large biomedical ontologies described in a typical medical language.

- Criteria: conformance with expected results, efficiency in terms of execution time.
- URL in SEALS testrepository http://seals.sti2.at/tdrs-web/testdata/persistent/Anatomy+Testsuite/2010/suite

hal-00794905, version 1 - 26 Feb 2013

Scenario 3: Conference The conference dataset consists of acollection of conference organization ontologies. This effort was expected to materialize in alignments as well as in interesting individual correspondences ('nuggets'), aggregated statistical observations and/or implicit design patterns.

- Criteria: conformance with expected results and alignment coherence.
- URL in SEALS testrepository: http://seals.sti2.at/tdrs-web/testdata/persistent/Conference+Testsuite/2010/suite

We are currently investigating the possibility of adding more scenarios to be executed in the SEALS modality. As soon as we know more, we will forward the information via a SEAL news article.

Timeline and Tutorial

More information can be found on the relevant OAEI pages. Note also that, due to supporting OAEI 2011, the timeline of the ontology matching campaign differs from the timeline of the other SEALS campaigns.

OAEI 2011 Webpage (http://oaei.ontologymatching.org/2011/) Informs about all tracks of OAEI 2011 and contains a timeline binding also for the tracks supported by SEALS. This page links also to track specific information.

SEALS platform evaluation modalities (http://oaei.ontologymatching.org/2011/seals-eval.html) Contains specific instructions for participants of tracks in SEALS modality. In particular, a tutorial for wrapping a matching tool and software that supports the process is offered for download. The concrete steps for participating are described in detail.

Tools have to be wrapped following the provided tutorial. Otherwise it is not possible to evaluate them in the SEALS platform.

How do I get involved?

If you want to participate, register your tool in the SEALS Portal (http://www.seals-project.eu) and stay tuned to the evaluation campaign web page where you can find detailed descriptions of the evaluations that we will perform and the latest information and results of the evaluation campaign.

If you have any question or comment about the evaluation campaign, please contact us.

- Jérôme Euzenat (*jerome.euzenat@inrialpes.fr*).
- Christian Meilicke (*christian@informatik.uni-mannheim.de*).
- Cássia Trojahn dos Santos (*cassia.trojahn-dos-santos@inrialpes.fr*).

We count on your participation!

hal-00794905, version 1 - 26 Feb 2013

C.5 Semantic Search Tools

C.5.1 Introduction

The goals of the semantic search tool evaluation initiative are to support developers to improve their tools; compare their tools against their competitors and to generally improve the interoperability of semantic technologies.

The short-term goal is to create a set of reference benchmark tests for assessing the strengths and weaknesses of the available tools and to compare them with each other. As such, these tests will focus on the performance of fundamental aspects of the tool in a strictly controlled environment / scenario rather than their ability to solve open-ended, real-life problems.

C.5.2 Criteria

Semantic search tools will be evaluated according to a number of different criteria including query expressiveness, usability (effectiveness, efficiency, satisfaction) and scalability. Scalability will address a number of factors including the tool's ability to query a large repository in a reasonable time; the tool's ability to cope with differing ontology sizes; and the tool's ability to cope with a large amount of query results. Query expressiveness will investigate the means by which queries are formulated within the tool and the degree to which this facilitates (or even impedes) the user's question-answering goal. However, given the interactive nature of semantic search tools, a core interest in this evaluation is the usability of a particular tool.

C.5.3 Two phase approach

The core functionality of a semantic search tool is to allow a user to discover one or more facts or documents by inputting some form of a query. The manner, in which this input occurs (e.g.: natural language, keywords, visual representation) is not of concern; however, the user experience of using the interface is of interest. Therefore, it is essential, that the evaluation procedures described in this document emphasize the users' experience with each tool.

In order to achieve this goal, the evaluation of each tool is split into two complementary phases: the automated phase and the user-in-the-loop phase. The user-in-the-loop phase involves a series of experiments involving human subjects, who are given a number of tasks (questions) to complete using a particular tool operating on an particular ontology.

Hence, the two core implications of this are that the user-in-the-loop experiments will be run by each tool provider participating in the evaluation and that additional software will be provided by this workpackage in order to both run the experimental workflows and likewise obtain the test data and return the results data to the various SEALS repositories. All materials required for the user-in-the-loop experiments will be provided by the SEALS consortium.

hal-00794905, version 1 - 26 Feb 2013

hal-00794905, version 1 - 26 Feb 2013

C.5.4 Take part

The evaluation campaign will take place by the end of 2011.

Participation is open to developers interested in evaluating their tool or to anyone who wants to evaluate a certain tool.

Participants are just expected to collaborate in the connection of their tool with the SEALS Platform, which will be the infrastructure that will run all the evaluations automatically. Besides checking their results and comparing with others, once the tool is connected to the SEALS Platform participants will also be able to run the evaluations on their own with these and future test data.

If you want to participate, register your tool in the SEALS Portal (http://www.seals-project.eu) and stay tuned to the evaluation campaign web page where you can find detailed descriptions of the evaluations that we will perform and the latest information and results of the evaluation campaign.

If you have any question or comment about the evaluation campaign, please contact us.

- Stuart Wrigley (*s.wrigley@dcs.shef.ac.uk*).
- Dorothee Reinhard (*dreinhard@ifi.uzh.ch*).

We count on your participation!

C.6 Semantic Web Service Tools

C.6.1 Evaluation Scenario Overview

The SWS discovery activity consists of finding Web Services based on their semantic descriptions. Tools for SWS discovery or matchmaking can be evaluated on retrieval performance, where for a given goal, i.e. a semantic description of a service request, and a given set of service descriptions, i.e. semantic descriptions of service offers, the tool returns the match degree between the goal and each service, and the platform measures the rate of matching correctness based on a number of metrics.

In the SEALS Semantic Web Service Discovery evaluation scenario we will test and compare SWS discovery tools/matchmakers on retrieval performance using the SEALS platform. SEALS services for applying standard IR measures such as Precision and Recall over registered datasets will be available. Retrieval results will be generated and stored for each registered tool.

C.6.2 Evaluation datasets

To evaluate the retrieval performance of SWS discovery tools/matchmakers, we will use:

- OWLS-TC
- SAWSDL-TC
- WSMO-LITE TC

C.6.3 Take part

The evaluation campaign will take place by the end of 2011.

Participation is open to developers interested in evaluating their tool or to anyone who wants to evaluate a certain tool.

Participants are just expected to collaborate in the connection of their tool with the SEALS Platform, which will be the infrastructure that will run all the evaluations automatically. Besides checking their results and comparing with others, once the tool is connected to the SEALS Platform participants will also be able to run the evaluations on their own with these and future test data.

If you want to participate, register your tool in the SEALS Portal (http://www.seals-project.eu) and stay tuned to the evaluation campaign web page where you can find detailed descriptions of the evaluations that we will perform and the latest information and results of the evaluation campaign.

If you have any question or comment about the evaluation campaign, please contact us.

- Liliana Cabral (*l.s.cabral@open.ac.uk*).
- Ning Li (*n.li@open.ac.uk*).
- Daniel Winkler (*daniel.winkler@sti2.at*).
- John Domingue (*j.b.domingue@open.ac.uk*).

We count on your participation!

hal-00794905, version 1 - 26 Feb 2013

D. Evaluation campaign agreements

The following agreements describe the general terms for participation in the SEALS Evaluation Campaigns and the policies for using the resources and results produced in these evaluation campaigns. These were defined in SEALS Deliverable 3.1 (García-Castro and Martín-Recuerda, 2009) and are reproduced here for completeness.

D.1 Terms of participation

By submitting a tool and/or its results to a SEALS Evaluation Campaign the participants grant their permission for the publication of the tool results on the SEALS web site and for their use for scientific purposes (e.g., as a basis for experiments).

In return, it is expected that the provenance of these results is correctly and duly acknowledged.

D.2 Use rights

In order to avoid any inadequate use of the data provided by the SEALS Evaluation Campaigns, we make clear the following rules of use of these data.

It is the responsibility of the user of the data to ensure that the authors of the results are properly acknowledged, unless these data are used in an anonymous aggregated way. In the case of participant results, an appropriate acknowledgement is the mention of this participant and a citation of a paper from the participants (e.g., the paper detailing their participation). The specific conditions under which the results have been produced should not be misrepresented (an explicit link to their source in the SEALS web site should be made).

These rules apply to any publication mentioning these results. In addition, specific rules below also apply to particular types of use of the data.

D.2.1 Rule applying to the non-public use of the data

Anyone can freely use the evaluations, test data and evaluation results for evaluating and improving their tools and methods.

D.2.2 Rules applying to evaluation campaign participants

The participants of some evaluation campaign can publish the results as long as they cite the source of the evaluations and in which evaluation campaign they were obtained.

Participants can compare their results with other published results on the SEALS web site as long as they also:

- compare with the results of all the participants of the same evaluation scenario; and
- compare with all the test data of this evaluation scenario.

hal-00794905, version 1 - 26 Feb 2013

Of course, participants can mention their participation in the evaluation campaign.

D.2.3 Rules applying to people who did not participate in an evaluation campaign

People who did not participate in an evaluation campaign can publish their results as long as they cite the sources of the evaluations and in which evaluation campaign they were obtained and they need to make clear that they did not participate in the official evaluation campaign.

They can compare their results with other published results on the SEALS web site as long as they:

- cite the source of the evaluations and in which evaluation campaign they were obtained;
- compare with the results of all the participants of the same evaluation scenario; and
- compare with all the test data of this evaluation scenario.

They cannot pretend having executed the evaluation in the same conditions as the participants. Furthermore, given that evaluation results change over time, it is not ethical to compare one tool against old results; one should always make comparisons with the state of the art.

D.2.4 Rules applying to other cases

Anyone can mention the evaluations and evaluation campaigns for discussing them.

Any other use of these evaluations and their results is not authorized (you can ask for permission however to the contact point) and failing to comply to the requirements above is considered as unethical.

hal-00794905, version 1 - 26 Feb 2013