

Semantic Evaluation at Large Scale Tutorial

Evaluating Semantic Web Services Tools

Liliana Cabral

Knowledge Media Institute
The Open University,
Milton Keynes, UK

l.s.cabral@open.ac.uk



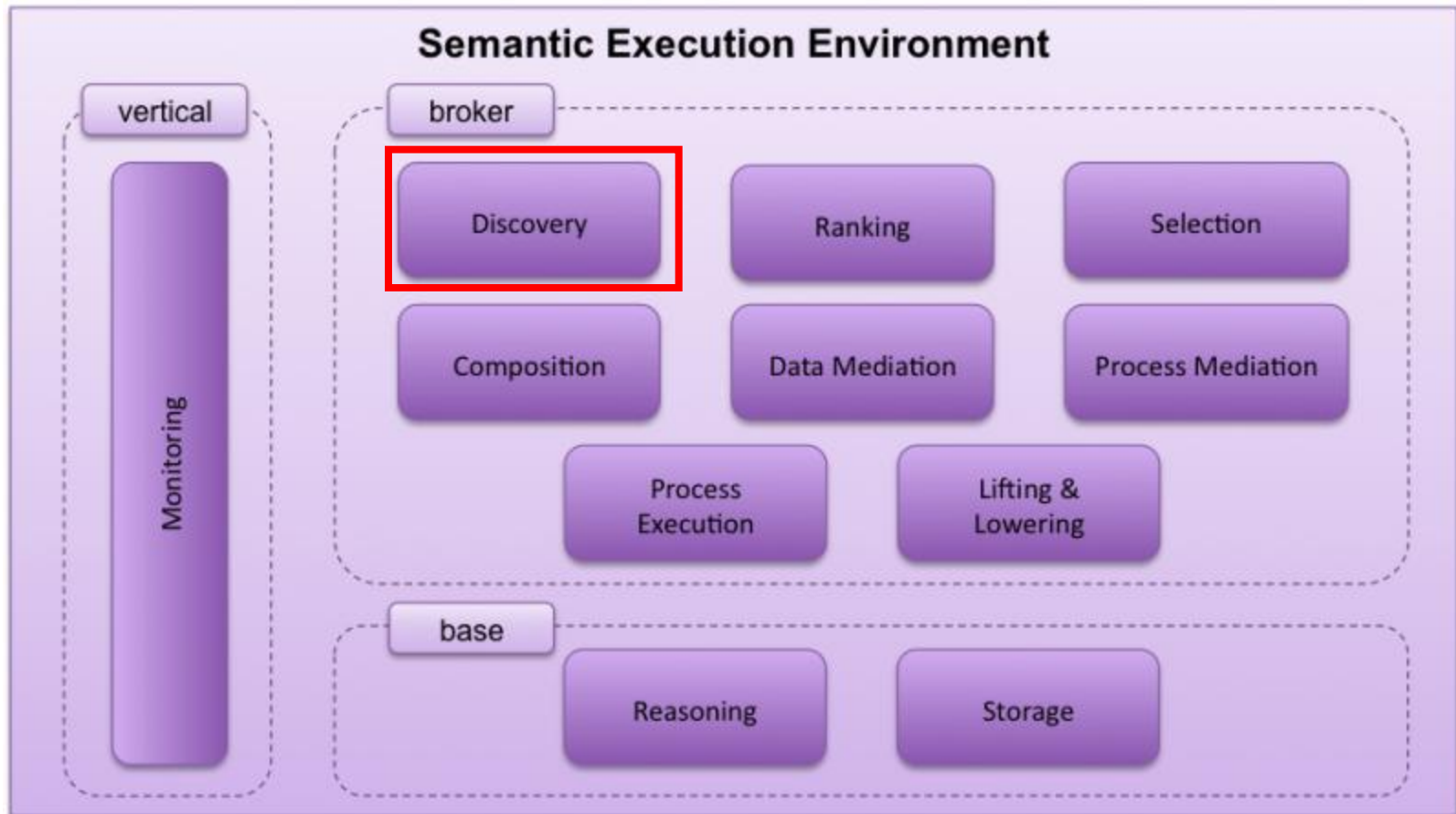
Outline

- Background
- Evaluation scenario
- Test data
- Results
- Evaluation Workflow
- Tools
- Conclusions

Background

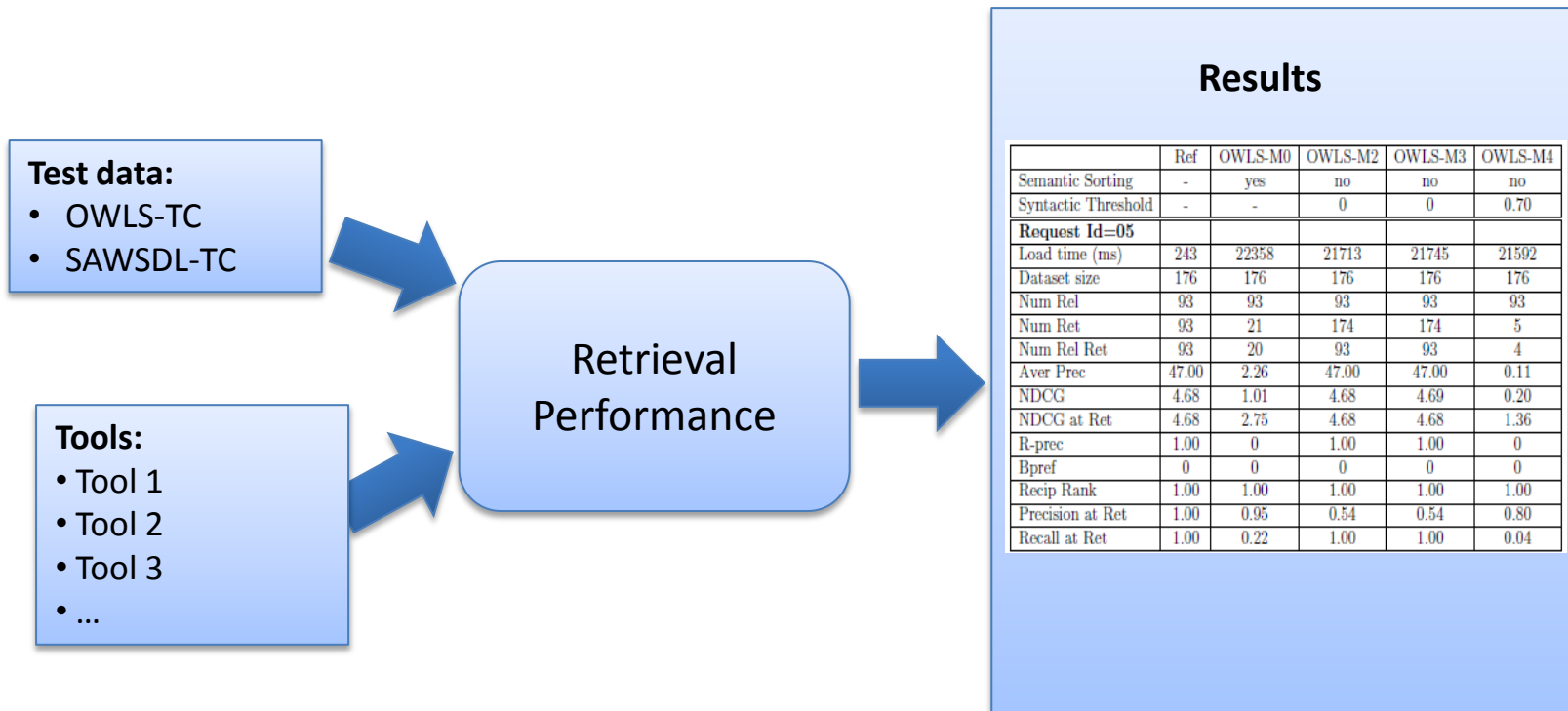
- Evaluation approach based on SWS standardization efforts from the OASIS Semantic Execution Environment Technical Committee (SEE-TC)
- The SEALS Approach is informed by existing SWS evaluation initiatives (SWS Challenge, S3, WSC)
- Uses existing test suites (e.g. OWLS-TC, SAWSDL-TC) and proposes new ones.

SEE architecture



Evaluation Scenario (2010)

<http://www.seals-project.eu/seals-evaluation-campaigns/semantic-web-services>



SWS Discovery Scenario

- The SWS discovery activity consists of finding Web Services based on their semantic descriptions.
- Tools for SWS discovery or matchmaking are evaluated on retrieval performance
- For a given goal, and a given set of service descriptions, the tool returns the match degree between the goal and each service,
- The platform provides measurement services for rating the matching correctness based on several metrics (e.g precision, recall)

Evaluation Services – Java APIs

- SWS Tool plugin API
 - Interface to be implemented by tool providers
 - Wrapper for running tool and providing raw results
- SWS Evaluation Measures API
 - Interface to be implemented with metrics (e.g. Precision, Recall)
 - Services for providing interpretations
- Repository Management API
 - Interfaces implemented for managing Testdata Repository and Results Repository
 - Adding and retrieving datasets and results
 - Parsing/Serialization to RDF

SWS Plugin API - Discovery

eu.sealsproject.domain.swst.plugin.activities

Interface Discovery

```
public interface Discovery
```

Method Summary

<u>DiscoveryResult</u>	<u>discover</u> (java.net.URI theGoalDescriptionURI)
void	<u>init</u> ()
void	<u>loadServices</u> (java.util.List<java.net.URI> services)

discover()

init()

loadServices()

Tool Bridge for SEALS platform

SWS Discovery interface methods	
<code>void init()</code>	Initializes the tool
<code>void loadServices(List serviceURIs)</code>	Load services for matchmaking
<code>DiscoveryResult discover(URI Goal)</code>	Invokes the tool. Returns the set of services that match the goal
ISemanticWebServiceDiscoveryToolBridge interface methods	
<code>boolean canExecute()</code>	Check if the tool bridge can be executed
<code>ToolType getType()</code>	Get the type of tool bridged by the implementation class

SWS Discovery Measures API

eu.sealsproject.domain.swst.evaluation.result
Class MetricsResult

```
java.lang.Object  
└─ eu.sealsproject.domain.swst.evaluation.result.MetricsResult
```

```
public class MetricsResult  
extends java.lang.Object
```

Constructor Summary

[MetricsResult\(\)](#)

Method Summary

double [getBpref\(\)](#)

double [getMap\(\)](#)

double [getNdcq\(\)](#)

double [getNdcq15\(\)](#)

double [getNum_rel_ret\(\)](#)

double [getNum_rel\(\)](#)

double [getNum_ret\(\)](#)

double [getPn\(\)](#)

double [getPrecision\(\)](#)

double [getR_prec\(\)](#)

double [getRecall\(\)](#)

double [getRecip_rank\(\)](#)

→ [getNumRetrieved\(\)](#)

→ [getNumRelevant\(\)](#)

→ [getAveragePrecision\(\)](#)

→ [getPrecisionAtCutoffPoint\(\)](#)

→ [getRecallAtCutoffPoint\(\)](#)

Available metrics by wrapping Galago*

Num Ret	Number of retrieved documents
Num Rel	Total number of documents judged relevant
Num Rel Ret	Number of retrieved documents that were judged relevant
Aver Prec	Average Precision. "Suppose the precision is evaluated once at the rank of each relevant document in the retrieval. If a document is not retrieved, we assume that it was retrieved at rank infinity. The mean of all these precision values is the average precision".
NDCG @ N	Normalized Discounted Cumulative Gain at cutoff point N. "This measure was introduced in Jarvelin, Kekalainen, "IR Evaluation Methods for Retrieving Highly Relevant Documents" SIGIR 2001. The formula is copied from Vassilvitskii, "Using Web-Graph Distance for Relevance Feedback in Web Search", SIGIR 2006. $Score = N \sum_i (2^{r(i)} - 1) / \log(1 + i)$, where N is such that the score cannot be greater than 1. We compute this by computing the DCG (unnormalized) of a perfect ranking".
NDCG	Normalized Discounted Cumulative Gain. "NDCG at (Math.max(retrieved.size(), judgments.size()))".
R-prec	R-Precision. "Returns the precision at the rank equal to the total number of relevant documents retrieved. This method is equivalent to precision(relevantDocuments().size())".

Bpref	Binary Preference. "The binary preference measure, as presented in Buckley, Voorhees "Retrieval Evaluation with Incomplete Information", SIGIR 2004. The formula is: $1/R \sum_r 1 - nrankedgreaterthanr /R$ where R is the number of relevant documents and n is a member of the set of first R judged irrelevant documents retrieved".
Recip Rank	Reciprocal Rank. "Returns the reciprocal of the rank of the first relevant document retrieved, or zero if no relevant documents were retrieved".
Precision @ N	Precision at cutoff point N. "Returns the precision of the retrieval at a given number of documents retrieved. The precision is the number of relevant documents retrieved divided by the total number of documents retrieved".
Recall @ N	Recall at cutoff point N. "Returns the recall of the retrieval at a given number of documents retrieved. The recall is the number of relevant documents retrieved divided by the total number of relevant documents for the query".

*<http://www.galagosearch.org/galagosearch-core/apidocs/>

Testdata

- For the 2010 Campaign we used the OWLS-TC 4.0 test collection (<http://projects.semwebcentral.org/projects/owls-tc/>).
- The OWLS-TC4.0 version consists of 1083 semantic web services described with OWL-S 1.1, covering nine application domains (education, medical care, food, travel, communication, economy, weapons, geography and simulation).
- OWLS-TC4 provides 42 test queries associated with binary as well as graded relevance sets. 160 services and 18 queries contain Precondition and/or Effects as part of their service descriptions.

Discovery Metadata – OWLS-TC

<http://seals.sti2.at/tdrs-web/testdata/persistent/OWLS-TC/4.0-7/suite/>

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:dc="http://purl.org/dc/terms/"

  xmlns:seals="http://www.seals-project.eu/ontologies/SEALSMetadata.owl#"
  xmlns:ds="http://www.seals-project.eu/ontologies/Discovery#">

  <rdf:Description rdf:about="http://www.seals-project.eu/Discovery#OWLS-TC_4">
    <rdf:type rdf:resource="http://www.seals-project.eu/ontologies/SEALSMetadata.owl
      #Suite" />
    <rdf:type rdf:resource="http://www.seals-project.eu/Discovery#DiscoverySuite" />
    <ds:hasDiscoveryTest rdf:resource="http://www.seals-project.eu/Discovery#
      DiscoveryTest1" />
    <ds:hasDiscoveryTest rdf:resource="http://www.seals-project.eu/Discovery#
      DiscoveryTest2" />
    <ds:hasDiscoveryTest rdf:resource="http://www.seals-project.eu/Discovery#
      DiscoveryTest3" />
    ...

    <seals:hasSuiteItem rdf:resource="http://www.seals-project.eu/Discovery#
      MatchTest_22_805"/>
    <seals:hasSuiteItem rdf:resource="http://www.seals-project.eu/Discovery#
      MatchTest_22_813"/>
    <seals:hasSuiteItem rdf:resource="http://www.seals-project.eu/Discovery#
      MatchTest_22_804"/>
    ...

  <rdf:Description rdf:about="http://www.seals-project.eu/Discovery#DiscoveryTest1">
    <rdf:type rdf:resource="http://www.seals-project.eu/Discovery#DiscoveryTest"/>
    <ds:hasMatchTest>http://www.seals-project.eu/Discovery#MatchTest_22_805</
      ds:hasMatchTest>
    <ds:hasMatchTest>http://www.seals-project.eu/Discovery#MatchTest_22_813</
      ds:hasMatchTest>
    <ds:hasMatchTest>http://www.seals-project.eu/Discovery#MatchTest_22_804</
      ds:hasMatchTest>
    <ds:hasMatchTest>http://www.seals-project.eu/Discovery#MatchTest_22_678</
      ds:hasMatchTest>
    <ds:hasMatchTest>http://www.seals-project.eu/Discovery#MatchTest_22_987</
      ds:hasMatchTest>
    ...
```

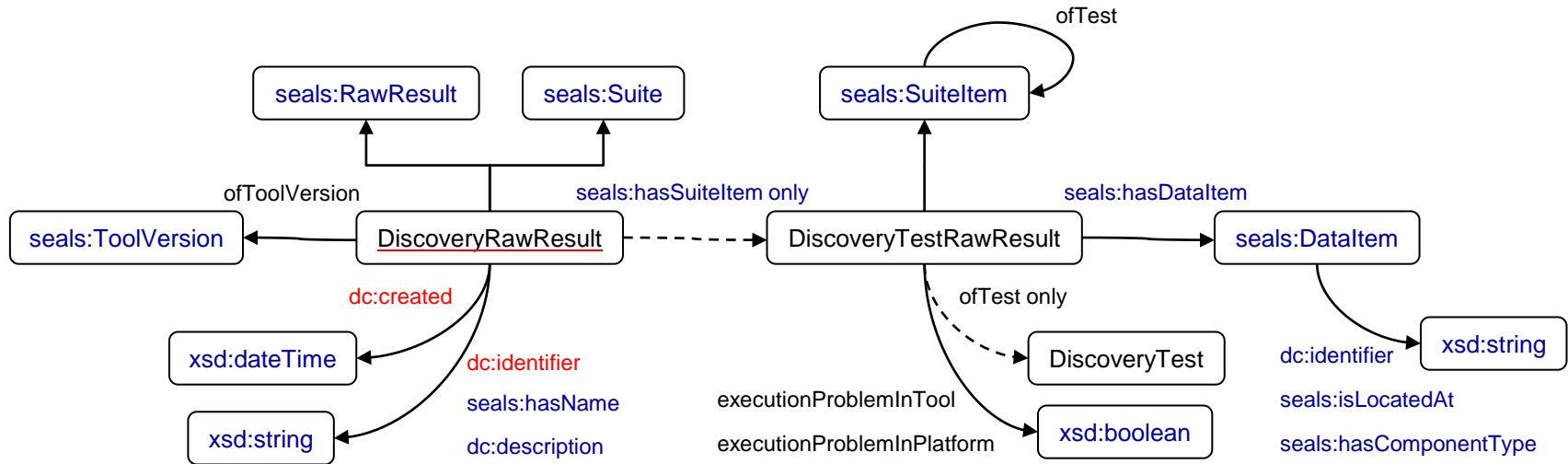
Discovery Ontology - Results

<http://www.seals-project.eu/ontologies/Discovery.rdf>

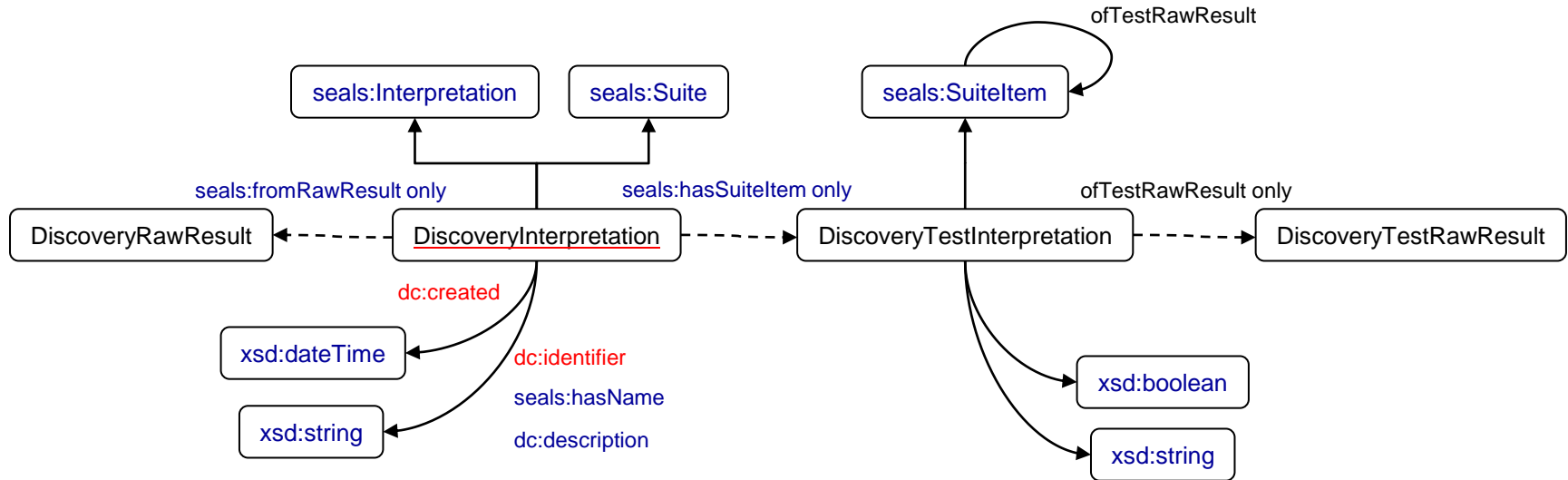
```
<rdf:RDF xmlns="http://www.seals-project.eu/Discovery#" xmlns:owl="http://www.w3.org/2002/07/owl#" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:smd="http://www.seals-project.eu/ontologies/SEALSMetadata.owl#" xmlns:xsd="http://www.w3.org/2001/XMLSchema#">
  <rdf:Description rdf:about="http://www.seals-project.eu/Discovery#MatchResult">
    <rdf:type>
      <rdf:Description rdf:about="http://www.w3.org/2000/01/rdf-schema#Class"/>
    </rdf:type>
    <rdfs:subClassOf>
      <rdf:Description rdf:about="http://www.seals-project.eu/ontologies/SEALSMetadata.owl#RawResult"/>
    </rdfs:subClassOf>
  </rdf:Description>
  <rdf:Description rdf:about="http://www.seals-project.eu/Discovery#DiscoveryResult">
    <rdf:type>
      <rdf:Description rdf:about="http://www.w3.org/2000/01/rdf-schema#Class"/>
    </rdf:type>
    <rdfs:subClassOf>
      <rdf:Description rdf:about="http://www.seals-project.eu/ontologies/SEALSMetadata.owl#RawResult"/>
    </rdfs:subClassOf>
  </rdf:Description>
  <rdf:Description rdf:about="http://www.seals-project.eu/Discovery#TestSuiteResult">
    <rdf:type>
      <rdf:Description rdf:about="http://www.w3.org/2000/01/rdf-schema#Class"/>
    </rdf:type>
    <rdfs:subClassOf>
      <rdf:Description rdf:about="http://www.seals-project.eu/ontologies/SEALSMetadata.owl#RawResult"/>
    </rdfs:subClassOf>
  </rdf:Description>
  ...

```

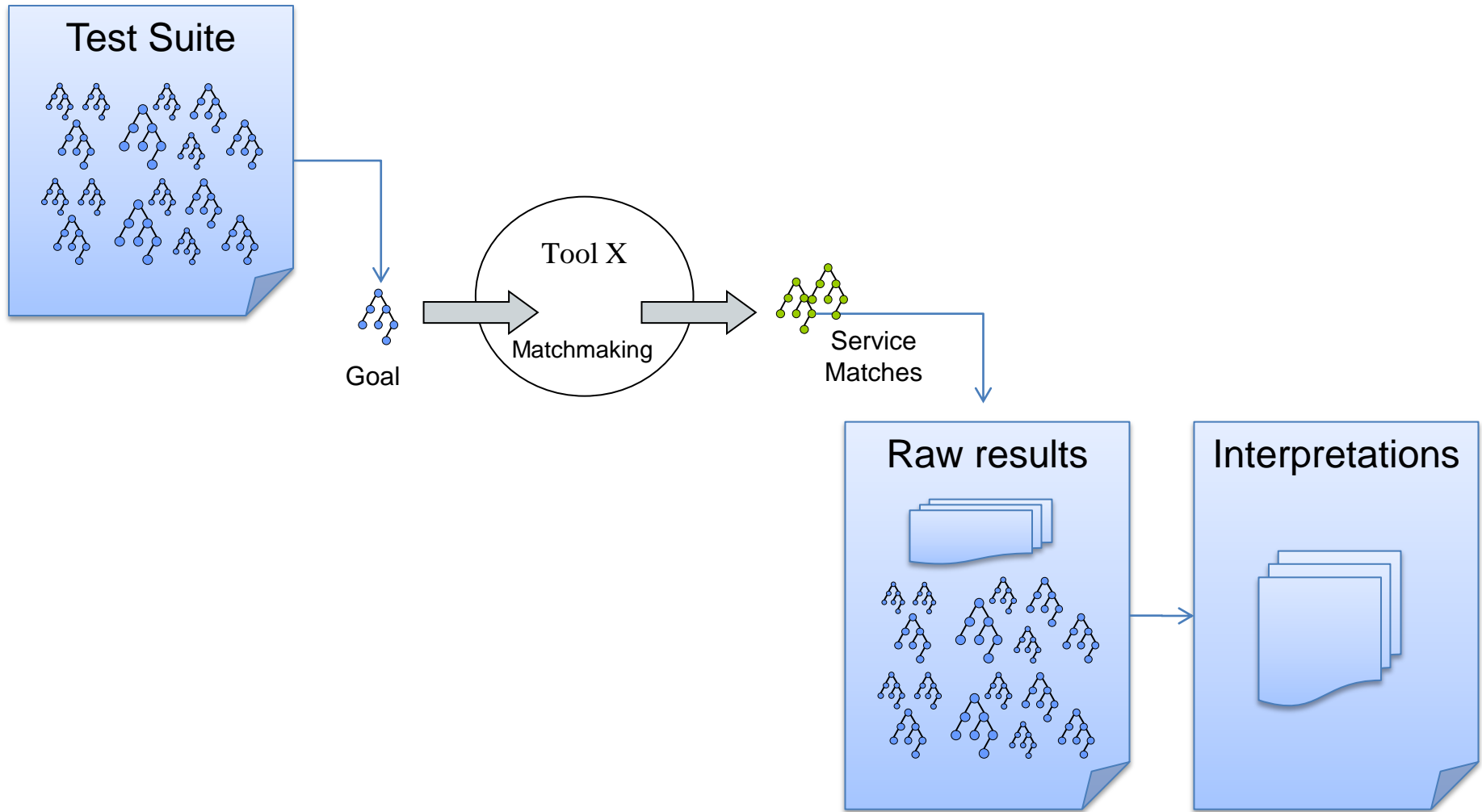
Describing Discovery Raw Results



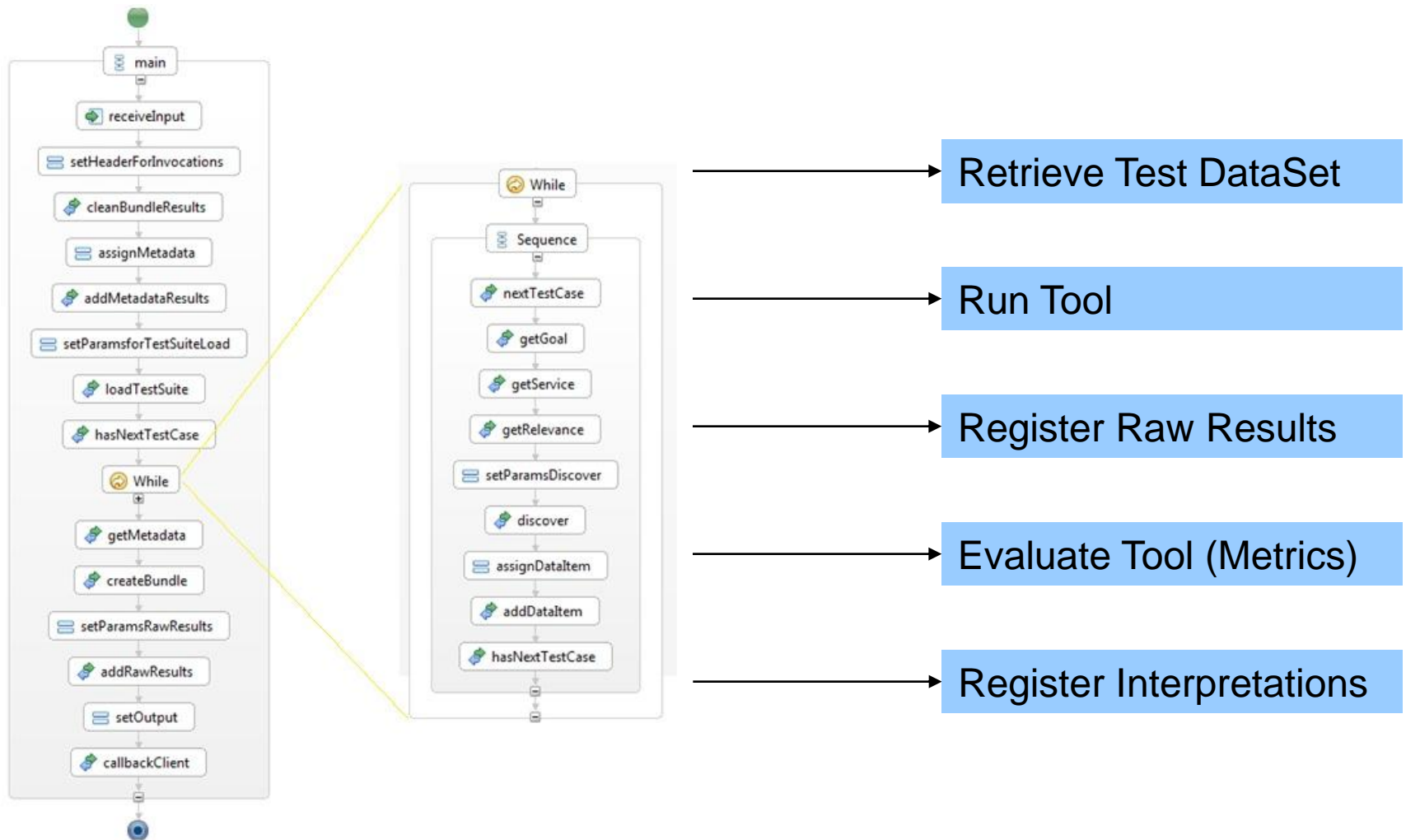
Describing Discovery Interpretations



Evaluation Workflow Overview



SWS Discovery Evaluation Workflow



Evaluation Campaign 2010

<http://www.seals-project.eu/seals-evaluation-campaigns/semantic-web-services>

- Experimental
- Organizers implemented the SWS plugin for OWLS-MX variants (different parameter settings)
- OWLS-TC4.0 (1083 service descriptions)
 - Datasets = Reference sets (size160-260)
- Results in SEALS Deliverable D14.3

Tools

Tool	Description
OWLS-M0	OWLS-MX variant - Constraint Similarity Measure
OWLS-M2	OWLS-MX variant - Extended Jaccard Similarity Measure
OWLS-M3	OWLS-MX variant - Cosine Similarity Measure
OWLS-M4	OWLS-MX variant - Jensen Shannon Similarity Measure

- OWLS-M0 – Semantic ordering; retrieved *exact, plugin, subsumes, subsumed-by* matches
- OWLS-M2 and OWLS-M3 – Syntactic ordering; threshold >0
- OWLS-M4 – Syntactic ordering; threshold >0.7

Goals (service requests)

Goal	Name	URI
request id="5"	car price service	http://127.0.0.1/queries/1.1/car_price_service.owl
request id="8"	2For 1 Price service	http://127.0.0.1/queries/1.1/dvd-playermp3player_price_service.owl
request id="21"	Researcher address service	http://127.0.0.1/queries/1.1/researcher-in-academia_address_service.owl

Evaluation Results (Example)

	Ref	OWLS-M0	OWLS-M2	OWLS-M3	OWLS-M4
Semantic Sorting	-	yes	no	no	no
Syntactic Threshold	-	-	0	0	0.70
Request Id=05					
Load time (ms)	243	22358	21713	21745	21592
Dataset size	176	176	176	176	176
Num Rel	93	93	93	93	93
Num Ret	93	21	174	174	5
Num Rel Ret	93	20	93	93	4
Aver Prec	47.00	2.26	47.00	47.00	0.11
NDCG	4.68	1.01	4.68	4.69	0.20
NDCG at Ret	4.68	2.75	4.68	4.68	1.36
R-prec	1.00	0	1.00	1.00	0
Bpref	0	0	0	0	0
Recip Rank	1.00	1.00	1.00	1.00	1.00
Precision at Ret	1.00	0.95	0.54	0.54	0.80
Recall at Ret	1.00	0.22	1.00	1.00	0.04

Conclusions

- The SWS plugin API is currently very similar to SME2's matchmaker plugin
- SEALS will extend to other activities
- Metrics from different sources can be added as services
- In the future other datasets will be available: WSMO-Lite
- Public intermediate results and repeatability is important for studying behaviour of the tool under different settings (not only 'best' behaviour)



Join the SEALS
Evaluation Campaigns!