

WS-DAI RDF(S) Realization

OGF20 DAIS Working Group

OGF IPR Policies Apply



- “I acknowledge that participation in this meeting is subject to the OGF Intellectual Property Policy.”
- Intellectual Property Notices Note Well: All statements related to the activities of the OGF and addressed to the OGF are subject to all provisions of Appendix B of GFD-C.1, which grants to the OGF and its participants certain licenses and rights in such statements. Such statements include verbal statements in OGF meetings, as well as written and electronic communications made at any time or place, which are addressed to:
 - the OGF plenary session,
 - any OGF working group or portion thereof,
 - the OGF Board of Directors, the GFSG, or any member thereof on behalf of the OGF,
 - the ADCOM, or any member thereof on behalf of the ADCOM,
 - any OGF mailing list, including any group list, or any other list functioning under OGF auspices,
 - the OGF Editor or the document authoring and review process
- Statements made outside of a OGF meeting, mailing list or other function, that are clearly not intended to be input to an OGF activity, group or function, are not subject to these provisions.
- Excerpt from Appendix B of GFD-C.1: “Where the OGF knows of rights, or claimed rights, the OGF secretariat shall attempt to obtain from the claimant of such rights, a written assurance that upon approval by the GFSG of the relevant OGF document(s), any party will be able to obtain the right to implement, use and distribute the technology or works when implementing, using or distributing technology based upon the specific specification(s) under openly specified, reasonable, non-discriminatory terms. The working group or research group proposing the use of the technology with respect to which the proprietary rights are claimed may assist the OGF secretariat in this effort. The results of this procedure shall not affect advancement of document, except that the GFSG may defer approval where a delay may facilitate the obtaining of such assurances. The results will, however, be recorded by the OGF Secretariat, and made available. The GFSG may also direct that a summary of the results be included in any GFD published containing the specification.”
- OGF Intellectual Property Policies are adapted from the IETF Intellectual Property Policies that support the Internet Standards Process.

Agenda

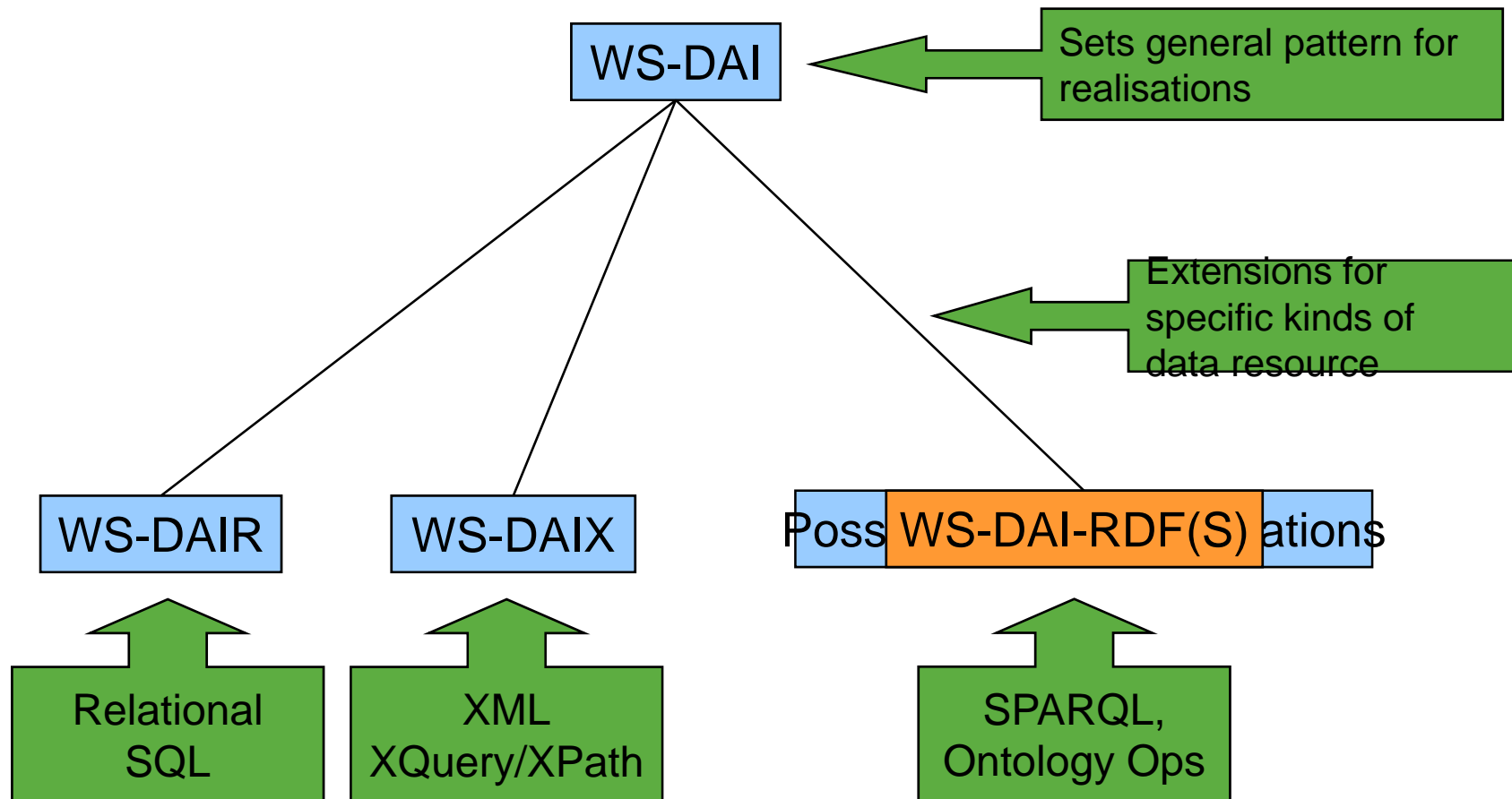
- WS-DAI-RDF(S) Introduction (Isao)
- Specification Introduction, Updates and Issues
 - *WS-DAI RDF(S) Querying* Specification
(Said and Isao)
 - *WS-DAI RDF(S) Ontology Access* Specification
(Oscar)
 - Glossary of Terms (Oscar)
- Discussion

WS-DAI RDF(S) Introduction

DAIS Working Group

Isao Kojima
*Grid Technology Research Center
AIST
Japan*

WS-DAI RDF(S) proposal



Motivation: Semantic Web and Grid Computing

(for other motivations, please read the motivational document)



Grid Computing: Beyond XML

- Grid Applications and Infrastructure will use Semantic Web technology
 - Ontology, Inference/Reasoning, etc

Example: Semantically enhanced Grid Resource Discovery

(presented later in this session)

Use of Ontology and Inference for discovering/monitoring grid resources

Resource description format: not XML, but RDF

Resource discovery language: not Xpath, but SPARQL

- **Semantic Web technology is mainly based on RDF-based format**
 - RDF, RDF Schema, OWL, OWL-S, SPARQL,,,,,
- **Current Grid standard is mostly based on XML-based format.**
 - XMLSchema, Xpath, XQuery etc.

Grid Infrastructure should support RDF-based format

If you want to use semantic web technology within grid applications

Semantic Web : Need to achieve Scalability

- Semantic Web applications and Tools should achieve Scalability as application grows.

Example: Large scale Distributed RDF data for ubiquitous ID tag

(also presented later in this session)

Each ID tag is associated with various semantic information described with RDF

- **Number of RDF data is huge and created in the distributed environment**
 - Distributed RDF databases
 - Terabyte scale of RDF storage

For large scale semantic/RDF applications,
Grid-based scalable & distributed RDF storage is necessary.

WS-DAI-RDF(S) Standard Structure

What is all about?

- *Providing an access mechanism to RDF(S) data resources*

New WS-DAI realization for RDF(S) databases
which will consist of 2 complementary specifications

1. RDF(S) Ontology Access

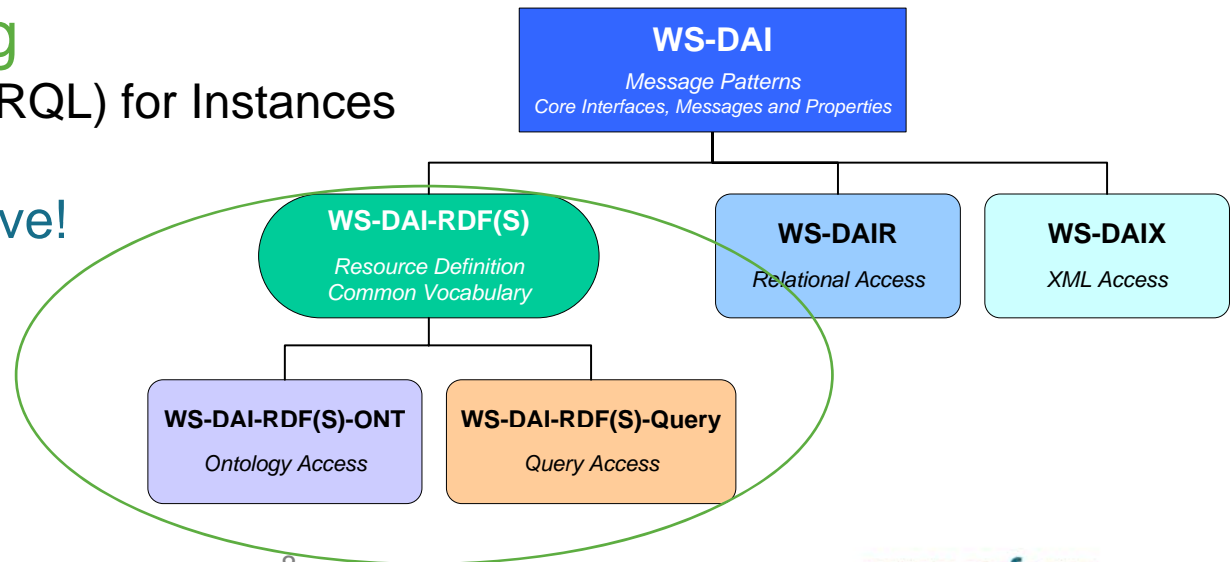
Ontological Primitives based on RDF(S) model(as Class)

2. RDF(S) Querying

Query Language(SPARQL) for Instances

Alternative, but not exclusive!

Use either or both
according to your needs



DAIS for RDF: History



- **2006.02: GGF16 at Athens**
 - DAIS for RDF BOF: Share the Motivation
- **2006.05: GGF17 at Tokyo**
 - Charter Discussion
 - Focus on RDF and RDF Schema (RDF(S))
 - Scope/Roadmap/Deliverables
- **2006.06: RDF F2F at Edinburgh**
 - Decide to make an informational document
 - Motivational Document : Structure Discussion
- **2006.09: GGF18 at Washington**
 - Motivational Doc Presented
- **2007.02 : OGF19 at Chapel Hill, NC**
 - 2 initial Specification Documents Presented

Current Documents



3 documents are on the forge

- Background & Motivational Scenarios

Initial Specs for

- Querying
- Ontology

Please download and have a look

DAIS WG
Open Grid Forum Working Group

Miguel Esteban Gutierrez, UH
Antonio Gomez-Perez, UH
Isao Kojima, AIS
Said Mirza Fakhri, AIS

Category: INFORMATIONAL July 28th, 2006

DAIS RDF(S) Realization: Background and Motivational Scenarios

Status of This Memo

This document provides information about the initiative for the provisioning of access to RDF(S) data resources by means of specific realizations of the WS-DAI Core specification.

Copyright Notice

Copyright © Global Grid Forum (2006). All Rights Reserved.

Abstract

The Database Access and Integration Services Working Group (DAIS-WG) has already submitted several specifications to the Global Grid Forum (GGF) recommendation track [WS-DAI, WS-DAIR, WS-DAIX]. These specifications provide a basic set of interfaces, properties and patterns for service-based data access within a grid. The core WS-DAI specification defines a set of common interfaces and operations that are independent of the type of underlying data resource, e.g. relational or XML data. X specifications respectively.

DAIS-WG This initiative is targeted towards and a proposal to extend the standard mechanism for accessing resources defined by the WS-DAI core

to that provide complementary ways logical-based primitives or by means

Web Services Data Access and Integration – The RDF(S) Realization (WS-DAIRDFS) RDF(S) Querying Specification, Version 0.1

Status of This Memo

This memo provides information regarding the specification of service-based interfaces to data resources. Distribution is unlimited.

Copyright Notice

Copyright © Open Grid Forum (2006). All Rights Reserved.

Abstract

Data resources play a significant role in many applications across multiple domains. Web services provide implementation neutral facilities for describing, invoking and orchestrating collections of networked resources. The OGF (Open Grid Forum) Open Grid Services Architecture (OGSA), and its associated specifications, defines consistent interfaces through web services to components of the grid infrastructure. Both the web and grid communities stand to benefit from the provision of consistent and agreed web service interfaces for data resources and the systems that manage them.

This document presents a specification for a collection of querying interfaces for RDF(S) data resources, which extends the interfaces defined in the Web Services Data Access and Integration document [WS-DAI]. It also presents interfaces for handling RDF graphs in RDF(S) data resources. This specification can be used with the complementary ontology access specification for RDF(S) data resources. This specification can be applied to regular web services environments or as part of a grid fabric.

Progress after OGF19

Progress after OGF19 is not large

- Revised version is not on the forge
- Periodical Telcons
 - Document Process Discussion
 - Need of “Common Glossary of Terminology”
To read both specifications
-> Will be discussed later
 - Need to have more “Use Cases”
To enhance motivational doc.
-> Some will be presented later.
 - How to increase the awareness of the activity
 - To get more Comments, Feedbacks, Interests, Use Cases, Volunteers, etc.,
- Specification Docs updates
 - -> Will be discussed here

Document Issues

- Background & Scenario Document
 - More Use Cases wanted
 - Grid based application for RDF databases
 - RDF Query & Ontology processing
- Specification Documents
 - **Ontology:** Still needs to be improved.
WSDL seems to be stable
 - **Querying:** Almost stable with WSDL (from AIST point of view)
- Any contributions are welcome
 - **Comments, Functional Requirements and Feedbacks** to the Documents

WS-DAI RDF(S) Querying

DAIS Working Group

Said Mirza Pahlevi & Isao Kojima
Grid Technology Research Center
AIST
Japan

Outline

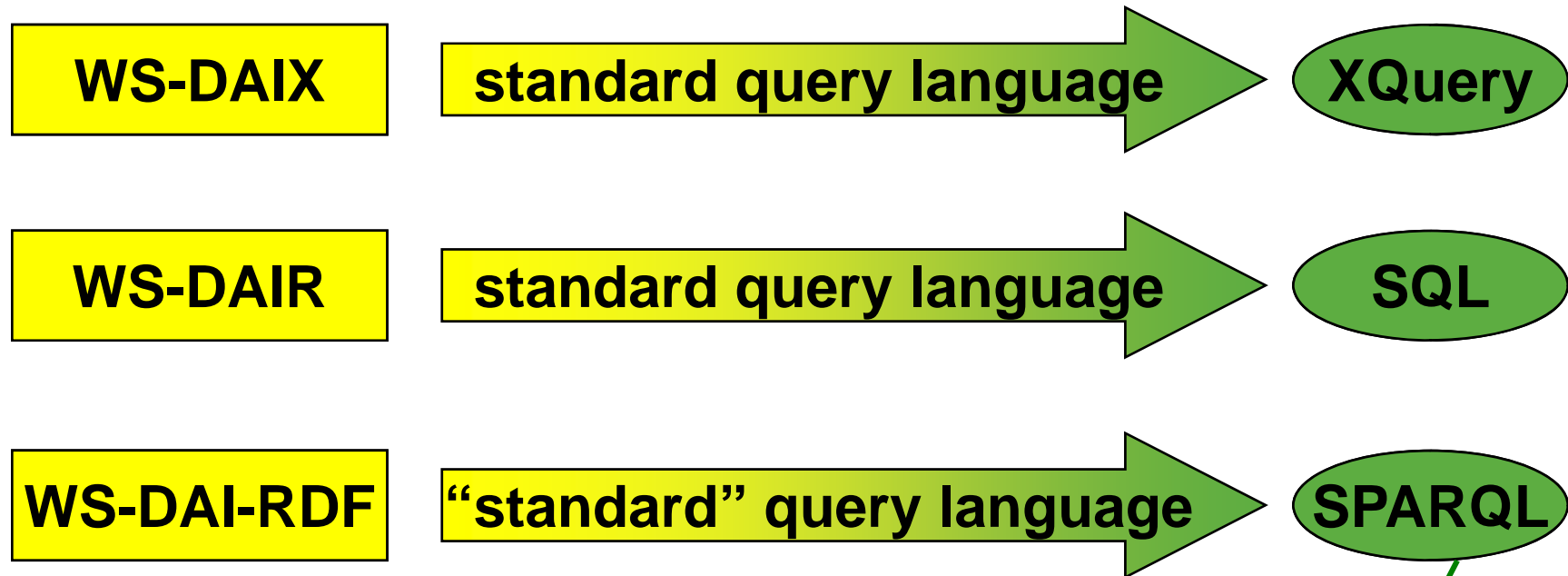
- Motivation & Aim of this presentation
 - Document Status & Update
- Document Specification Introduction
- Current Use Cases
- Issues

Motivation & Document Status

WS-DAI RDF(S) Querying

- Why do we need the query specification?
 - There are many Semantic Grid projects
 - The projects handle a large amount of RDF data
 - The need to access and integrate the data in a standard grid-based interface with some QUERY LANGUAGE.
- Why not only W3C SPARQL specifications?
 - Query language, results XML Format & **protocol**
 - **No grid specific functions**
 - No factory access pattern
 - No stateful Web service concept
 - RDF data collection management
- Extend the DAIS spec. by incorporating some parts of the SPARQL specifications

SPARQL as a Query Language



- W3C working draft
- More than 20 query engine implementations (<http://esw.w3.org/topic/SparqlImplementations>)

WS-DAI RDF(S) Querying (No change from GGF19)



- **Feature:**

- Focus to Instances, Not Schema.
 - *RDF Graph (Triples)*
- Focus to Query Language, Not API.
 - *W3C SPARQL*

- **Design Principle:**

Minimum extension /modification

to the existing W3C standards and WS-DAI core

- Supports W3C standards
 - *W3C SPARQL*
 - *W3C SPARQL result XML format (and RDF/XML,)*
 - *W3C SPARQL protocol for RDF (as possible)*
- Supports WS-DAI core model.
 - *Direct and Indirect Access*
- Supports Some Useful Functions
 - *Graph operations (including some update functions)*

Web Services Data Access and Integration – The RDF(S) Realization
(WS-DAIRDFS) RDF(S) Querying Specification, Version 0.1

[Status of This Memo](#)

This memo provides information regarding the specification of service-based interfaces to data resources. Distribution is unlimited.

[Copyright Notice](#)

Copyright © Open Grid Forum (2006). All Rights Reserved.

Abstract

Data resources play a significant role in many applications across multiple domains. Web services provide implementation neutral facilities for describing, invoking and orchestrating collections of networked resources. The OGF (Open Grid Forum) Open Grid Services Architecture (OGSA), and its associated specifications, defines consistent interfaces through web services to components of the grid infrastructure. Both the web and grid communities stand to benefit from the provision of consistent and agreed web service interfaces for data resources and the systems that manage them.

This document presents a specification for a collection of querying interfaces for RDF(S) data resources, which extends the interfaces defined in the Web Services Data Access and Integration document [WS-DAI]. It also presents interfaces for handling RDF graphs in RDF(S) data resources. This specification can be used with the complementary ontology access specification for RDF(S) data resources. This specification can be applied to regular web services environments or as part of a grid fabric.

Table of Contents (1)

(No change from OGF19)

- 1 Introduction
- 2 Notational Conventions
- 3 Terminology
 - 3.1 RDF(S) Data Resource
 - 3.2 RDF(S) Interfaces
- 4 RDF(S) Collection
 - 4.1 Static RDF(S) Collection Description
 - 4.1.1 RDFSCollection
 - 4.1.2 NumberOfGraphs
 - 4.2 Configurable RDF(S) Collection Description
 - 4.3 Example of RDFSCollectionPropertyDocument
- 4.4 RDFSCollectionAccess
 - 4.4.1 RDFSCollectionAccess::GetCollectionPropertyDocument
 - 4.4.2 RDFSCollectionAccess::AddGraphs
 - 4.4.3 RDFSCollectionAccess::GetGraphs
 - 4.4.4 RDFSCollectionAccess::RemoveGraphs
- 4.5 RDFSCollectionFactory
 - 4.5.1 RDFSCollectionFactory::GraphSelectionFactory

RDF(S)
Data Resource

Graph
Operations

Table of Contents (2)

(No change from GGF19)

5 SPARQL

5.1 Static SPARQL Description

5.5 SPARQLAccess

5.5.1 SPARQLAccess::GetSPARQLPropertyDocument

5.5.2 SPARQLAccess::SPARQLExecute

SPARQL
Access

5.6 SPARQLFactory

5.6.1 SPARQLFactory::SPARQLExecuteFactory

6 SPARQLItemsSet

6.1 Static SPARQLItemsSet Description

6.1.1 NumberOfItems

Description of
Query Results

6.5 SPARQLResultSetAccess

6.5.1 SPARQLResultSetAccess::GetSPARQLItemsSetPropertyDocument

6.5.2 SPARQLResultSetAccess::GetResults

ResultSet
Access

6.6 SPARQLTriplesSetAccess

6.6.1 SPARQLTriplesSetAccess::GetSPARQLItemsSetPropertyDocument

6.6.2 SPARQLTriplesSetAccess::GetTriples

Triples
Access

WSDLs

- Presented in the Appendix of the doc.
- Almost stable as an initial version

Document Status Update

Only one major change from OGF19

- GraphName(string) -> GraphNameURI(URI) (type change)
This will be discussed in the “Issues” section
- However, we could not get so many feedbacks and comments.
 - At least the doc skeleton structure seems to be OK from AIST point of view.
 - We think we need to have as many feedbacks to this initial spec.
- Working Item: We will attach the another Appendix sections including
 - Use Cases Section
 - Issues Section
- These sections are not included in the final specification document, however, we think it will be useful to appeal to a wider audiences.
 - as advised at Telcon

Querying Specification Overview

WS-DAI RDF(S) Querying

Basically, the spec is unchanged from OGF19

RDF(S) Data Resource for Querying



(Definition of Collection is an issue)

Data Resource=RDFSCollection = TopLevelCollection of WS-DAIX

- No Hierarchical Structure
- Holds a set of Graphs (=RDF DataSet)
- Each graph is uniquely identified by URI

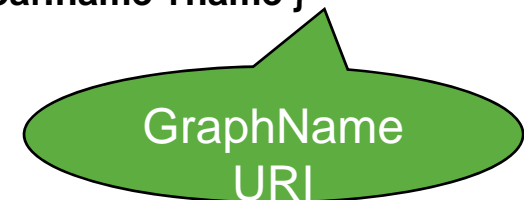
SPARQL requires **Graph**

- **Graph** is (based on W3C definition)
 - A set of Triples
 - Identified by URI

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?name
FROM <http://example.org/foaf/aliceFoaf>
WHERE { ?x foaf:name ?name }
```

RDF DataSet is defined in SPARQL specification

- **RDF DataSet** is
 - A set of **Graphs** including;
 - One (or many?) default graph
 - Zero or more NAMED graphs



Graph Operations (issue)



Basic Graph Operations is supported

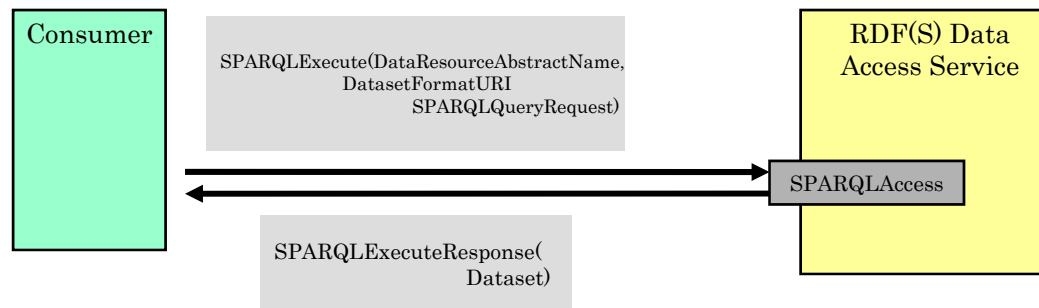
- **AddGraphs**
- **RemoveGraphs**
- **GetGraphs**

Not standard at all, but these definitions are similar with Document operations of WS-DAIX

- Seems to be useful to support
- **GetPropertyDocument(to get a list of graphs)**
- **GraphSelectionFactory**

SPARQL Interface

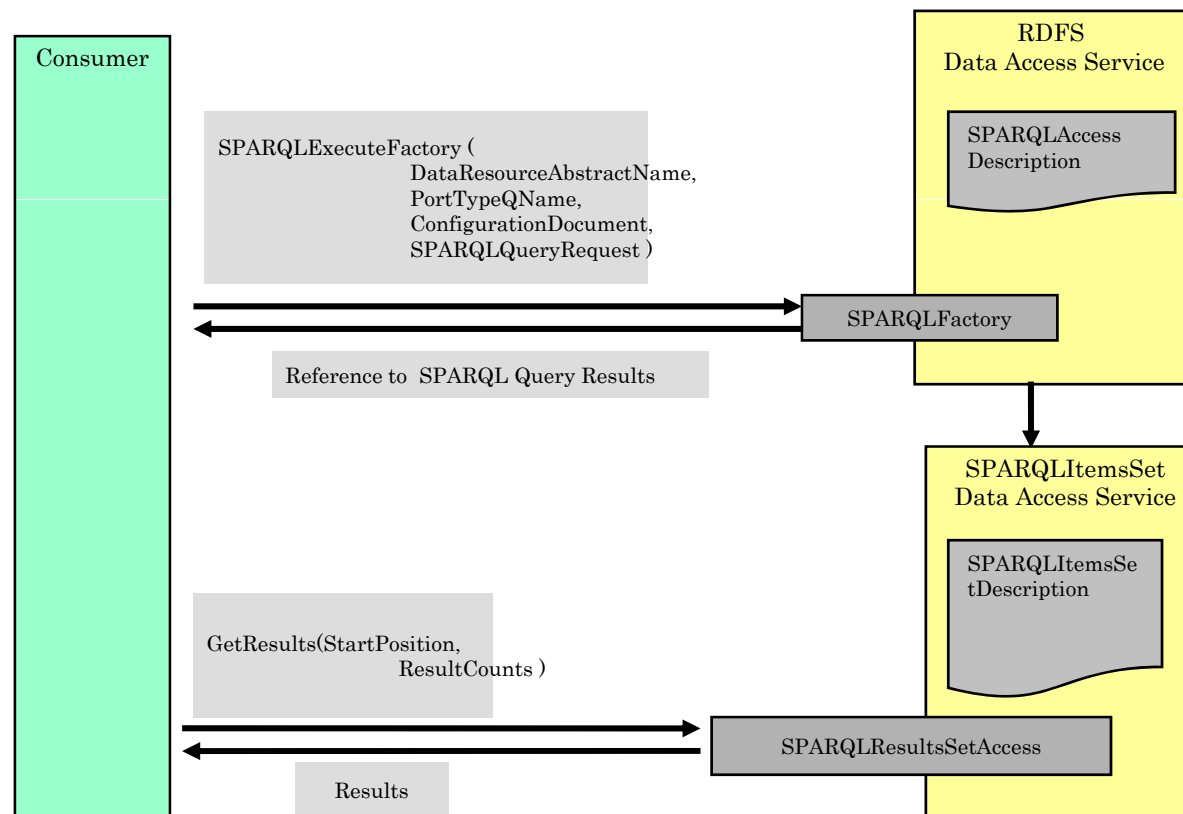
Same Pattern with WS-DAIR(SQL) and DAIX(Xpath/XQuery)



- **Input:**
 - DataResourceAbstractName
 - DatasetFormatURI
 - SPARQLQueryRequest
- **Output:**
 - Dataset with
 - W3C SPARQL QueryResult Format, or
 - RDF Format
- **SPARQL Description**
 - ExternalGraphAccess

Indirect Access

Similar structure with WS-DAIR, DAIX



Difference with WS-DAIR, DAIX



SPARQL returns

- A set of triples(=graph) for Construct/Describe
- A set of bindings for Select

Similar structure with Rowset and XMLSequence of WS-DAIR/X

Difference

- 2 types of data format => 2 Interfaces
- One Data Description (ItemsSet)

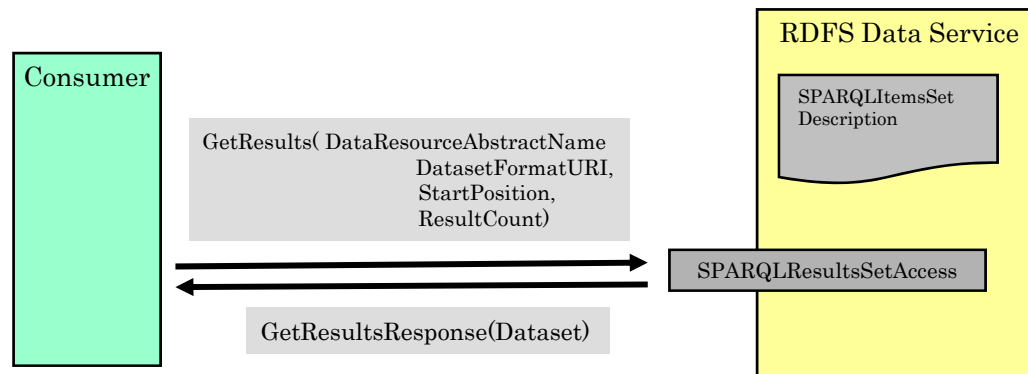
Common Data Description

- SPARQLItemsSet (for generic query results)
 - Number of Items(Number of bindings or Number of Triples)

Access Interfaces (This is an issue whether we should have two)

- ResultsSet Interface for Select/Ask(SPARQL/XML Format)
 - Get one(or many) binding(s) from a result Set
- TriplesSet Interface for Construct/Describe(RDF Format)
 - Get one(or many) Triple(s) from a result graph

SPARQL ResultSet Access

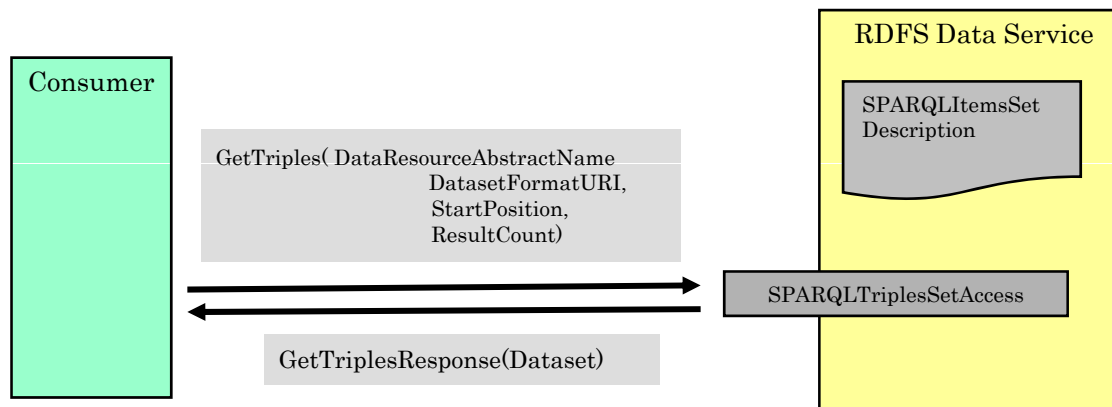


```
<?xml version="1.0"?>
<sparql xmlns="http://www.w3.org/2005/sparql-results#">
<head> ....</head>
<results ordered="false" distinct="false">
  <result>
    <binding name="x"> ... </binding>
    <binding name="hpage"> ... </binding>
  </result> ..
  <result>
    <binding name="x"> ... </binding>
    <binding name="hpage"> ... </binding>
  </result>
</results>
</sparql>
```

Startposition
&
ResultCount

SPARQL TripleSetsAccess

- Same Structure



- Get triple(s) from the graph

Use Case Examples

WS-DAI RDF(S) Querying

Outline

- Semantic matchmaking of Grid resources
- RDF data snapshot
- Large scale distributed RDF storage
- Federated SPARQL query processing

Semantic Matchmaking of Grid Resources

Grid Resource Matching



- **Motivation:**

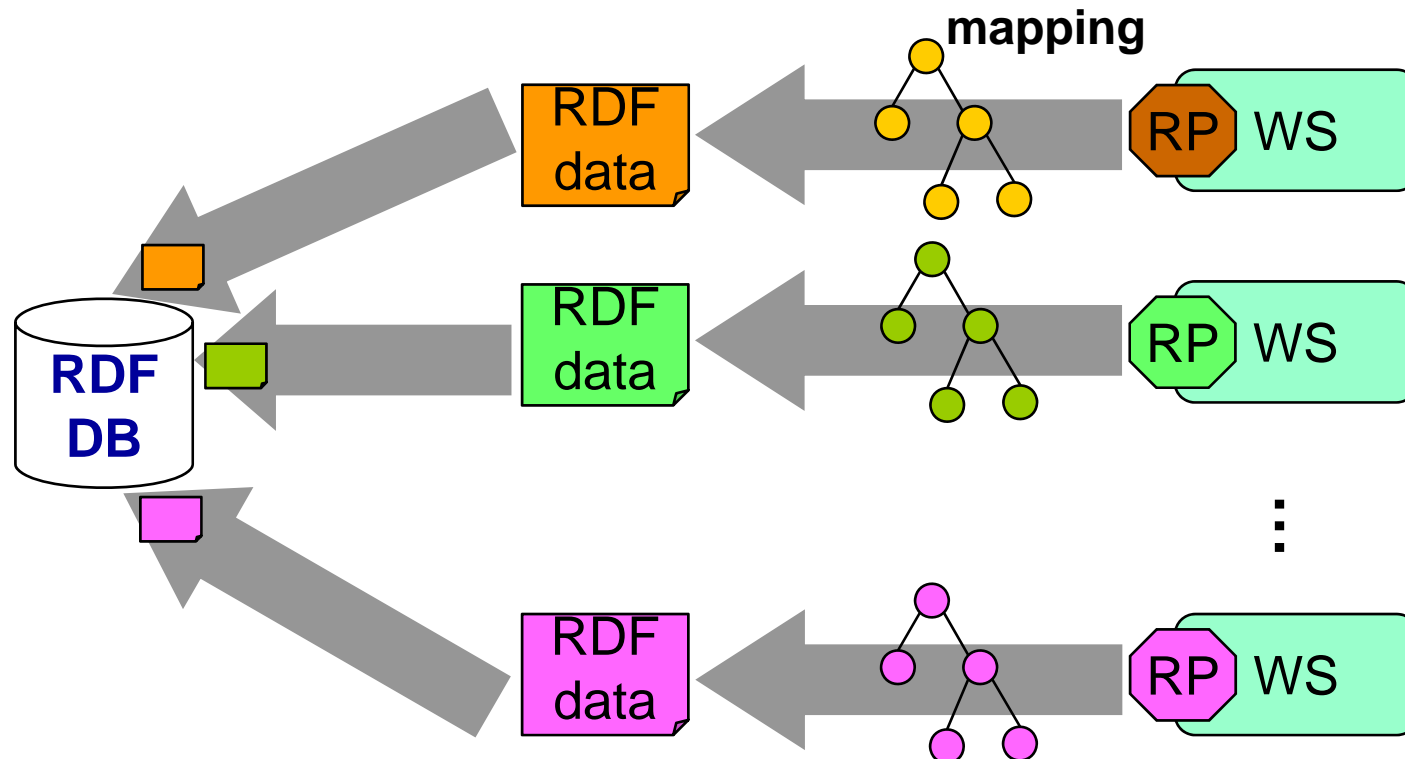
- Works describing Grid resources/services using RDF/OWL
 - S-MDS, *Said et al.*, OMM, *Tangmunarunkit et al.*
- The need for accessing various RDF-based resource descriptions and matching resources

- **Goal:**

- Semantic Grid resource matching using SPARQL

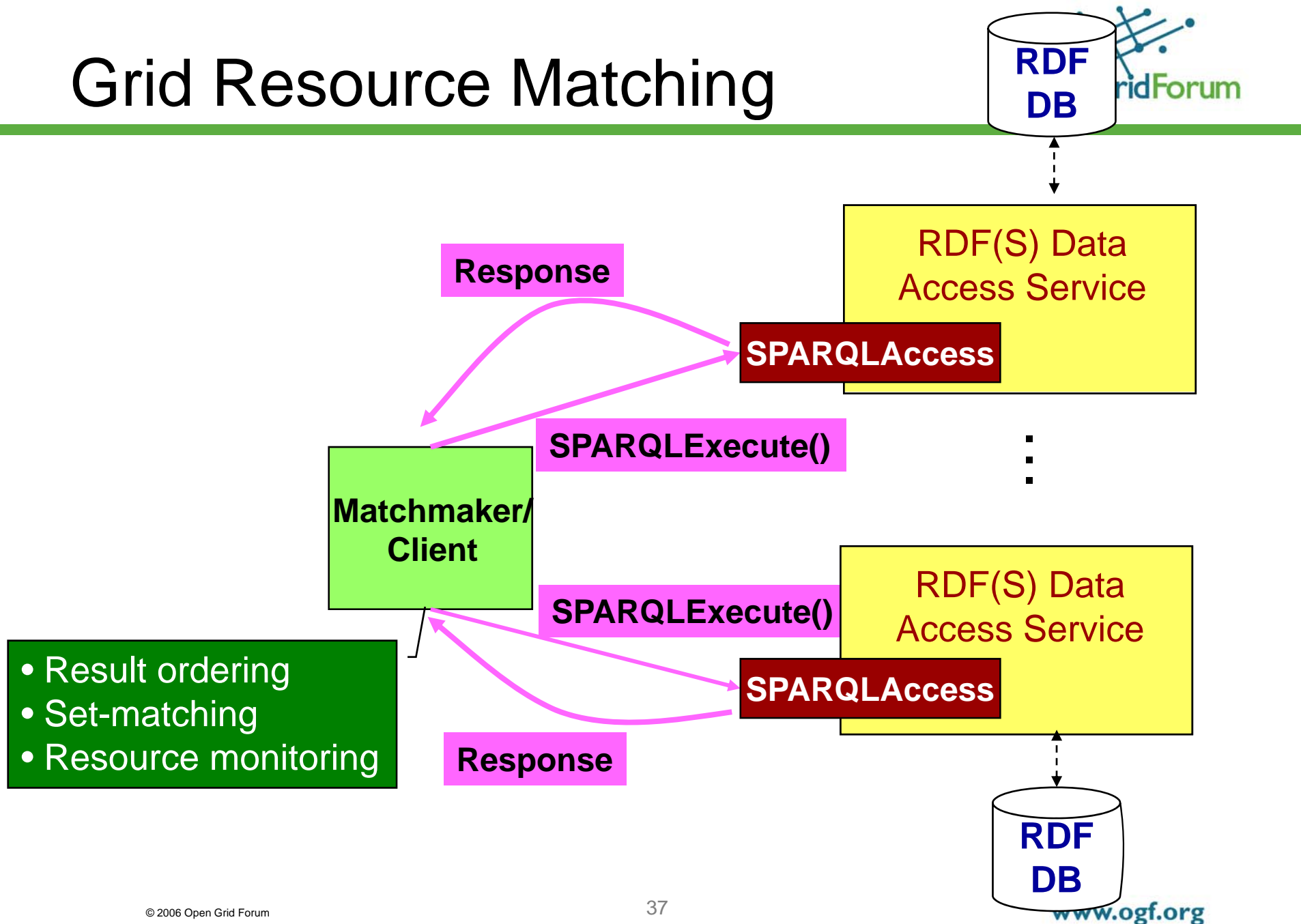
S-MDS

Provides a mechanism to **map a resource property (RP) into ontology** and to **aggregate and maintain the mapping results**



```
<Host Name="dbgrid.x.x" UniqueID="dbgrid.x.x">  
  <Processor CacheL1="0" CacheL1D="0" CacheL1I="0"  
    CacheL2="0" ClockSpeed="3001" InstructionSet="x86"/>  
  <MainMemory RAMAvailable="19" RAMSize="1484"  
    VirtualAvailable="332" VirtualSize="2041"/>  
  <OperatingSystem Name="Linux"  
    Release="2.6.0-test5_2smp"/>  
  <Architecture SMPSize="2"/>  
  <FileSystem AvailableSpace="65511" Name="entire"  
    ReadOnly="false" Root="/" Size="82459"/>  
  <NetworkAdapter IPAddress="163.220.2.54"  
    InboundIP="true" MTU="0" Name="dbgrid.hpcc.jp"  
    OutboundIP="true"/>  
  <ProcessorLoad Last15Min="10" Last1Min="6"  
    Last5Min="8"/>  
</Host>
```

Grid Resource Matching

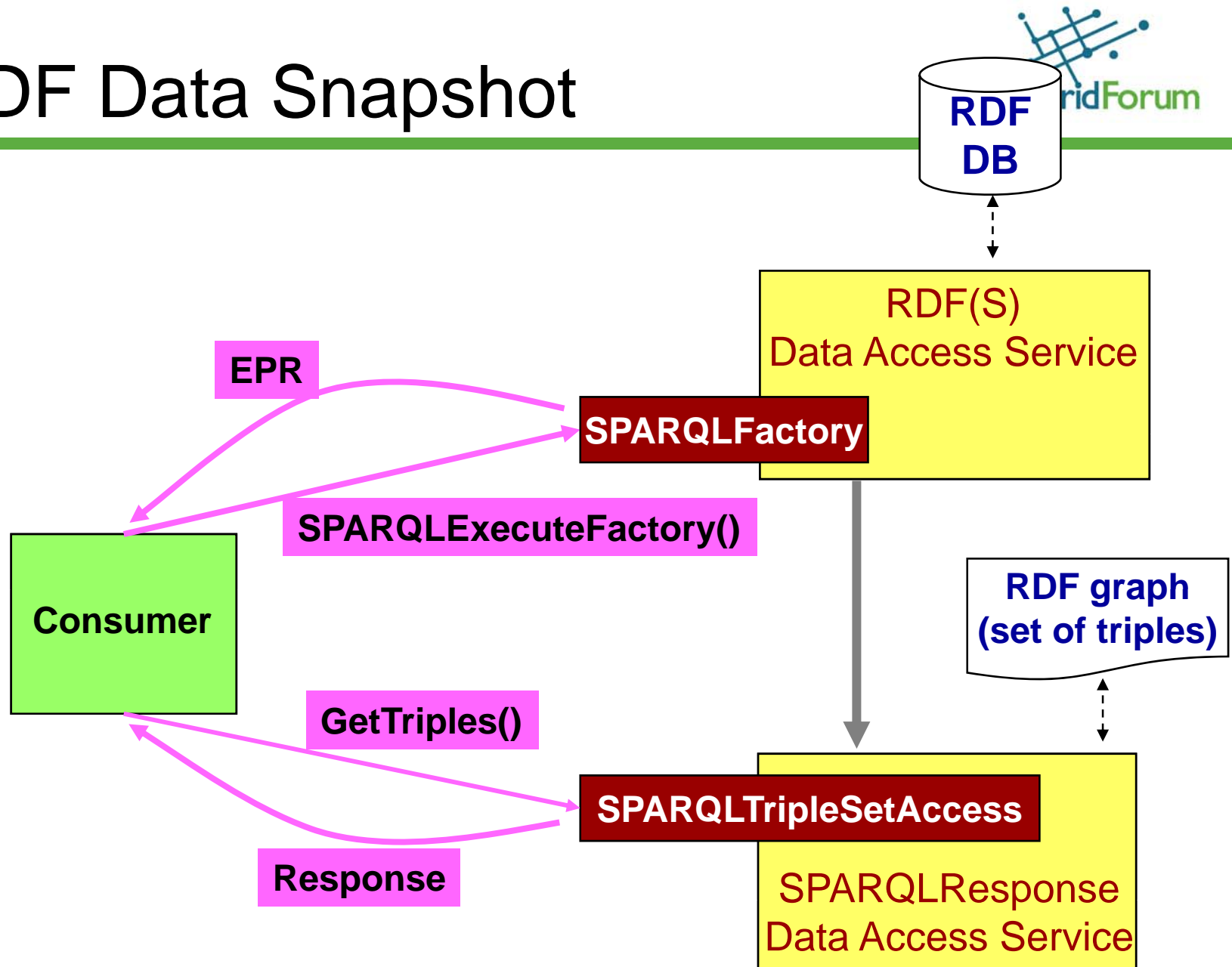


RDF Data Snapshot

RDF Data Snapshot

- Motivation:
 - Dynamic RDF data
 - Processor load and available memory
 - Road traffic and weather information
 - The need to obtain a data snapshot of a certain time point
- Goal:
 - Providing an RDF data snapshot

RDF Data Snapshot



Large Scale Distributed RDF Storage

Large Scale Distributed RDF Storage

- Motivation

- Ucode: Ubiquitous Code

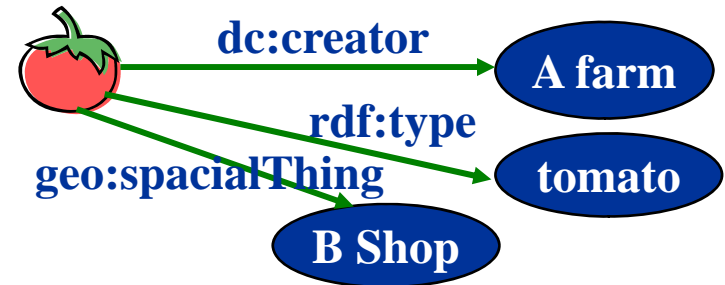
- Has 128 bit code for identifying ubiquitous object
 - Uses in RFID tag applications
 - **Has RDF as the information model** (DC, FOAF etc.)

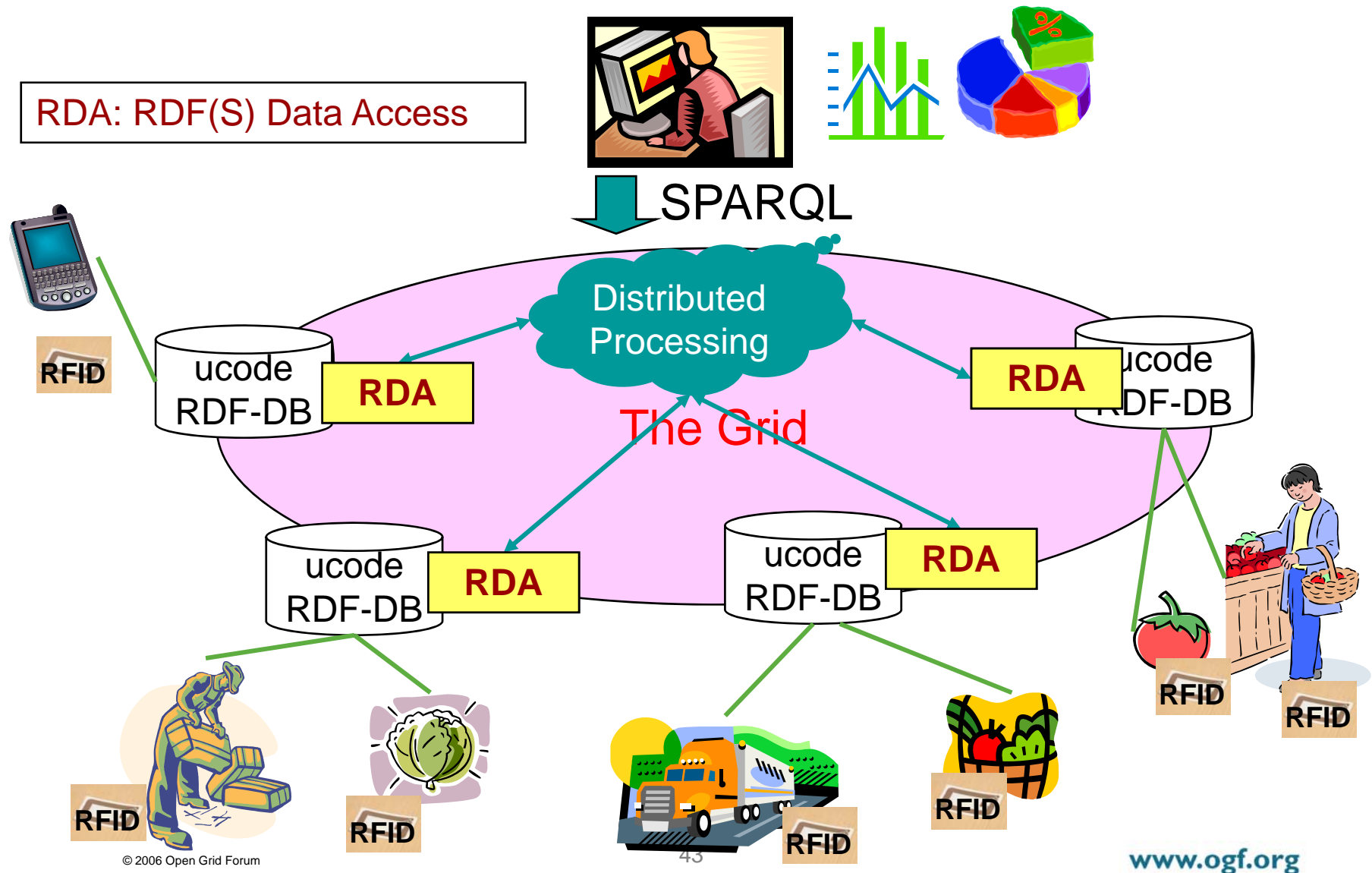
- A huge distributed RDF data

- SPARQL for UCode resolution

- Goal

- Provide a scalable distributed RDF storage that supports SPARQL





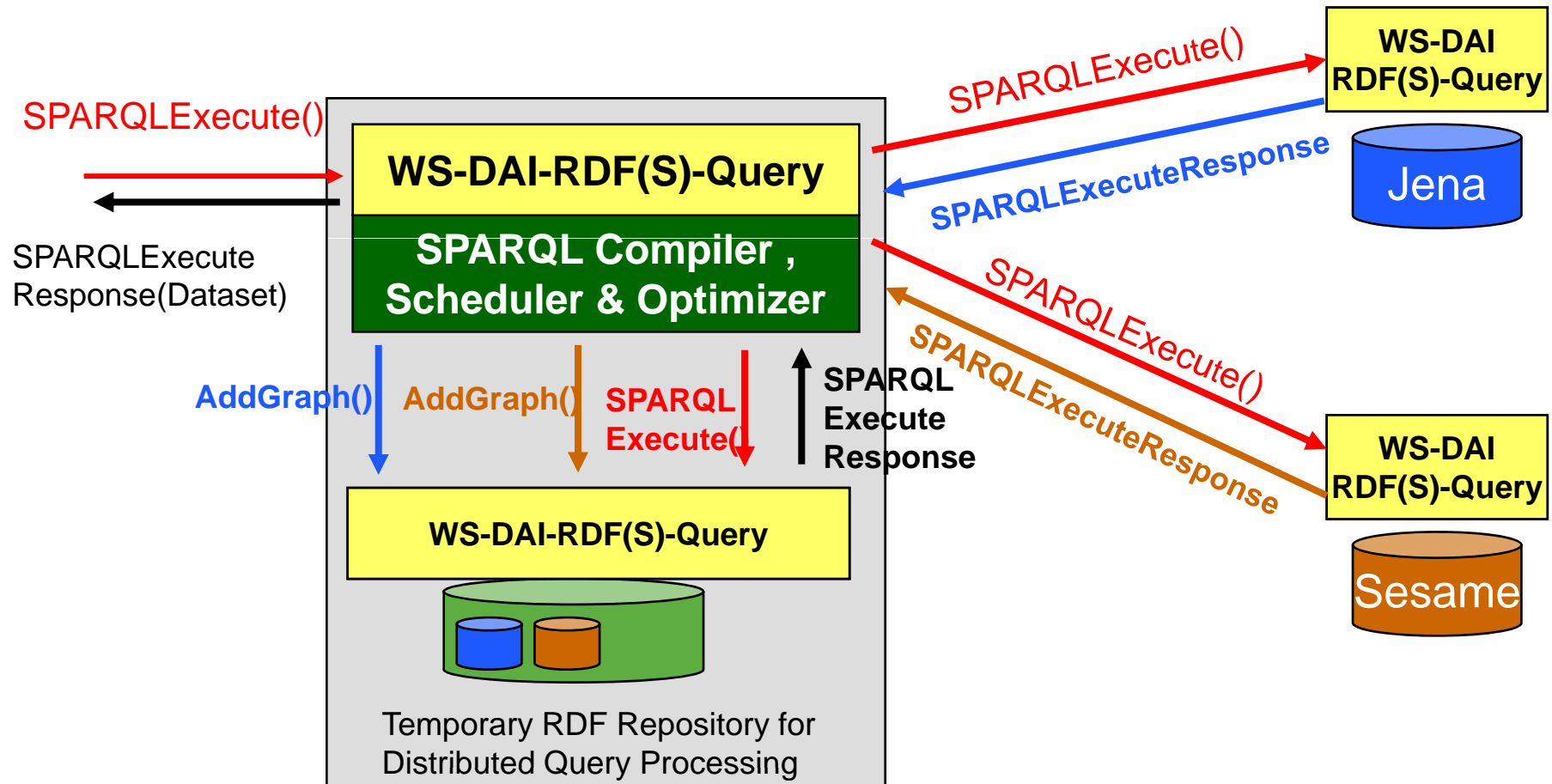
Federated SPARQL Query Processing

Federated SPARQL Query Processing



- Motivation
 - RDF data is created in bottom-up manner
 - Distributed RDF repositories
 - Managed by an RDF software/tool that supports SPARQL (e.g., Jena, Sesame, and Boca)
 - Heterogeneous Platform
 - Data in different repositories may be relevant to each other
 - FOAF
- Goal
 - Enabling SPARQL query over distributed RDF repositories that joins the RDF data

Federated SPARQL Query Processing

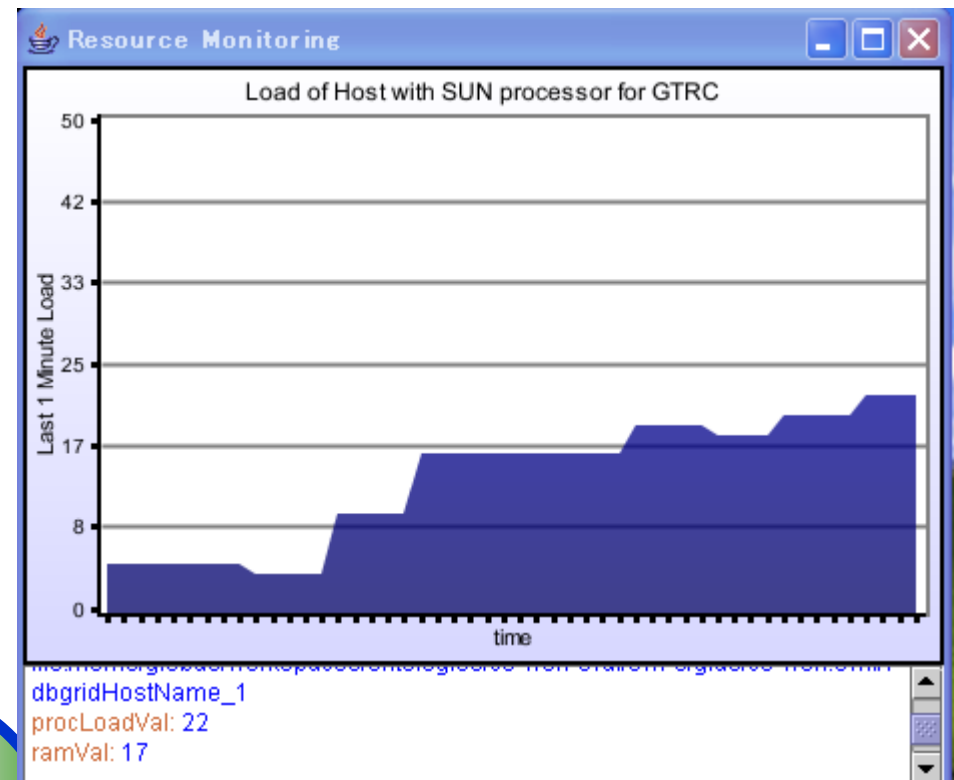
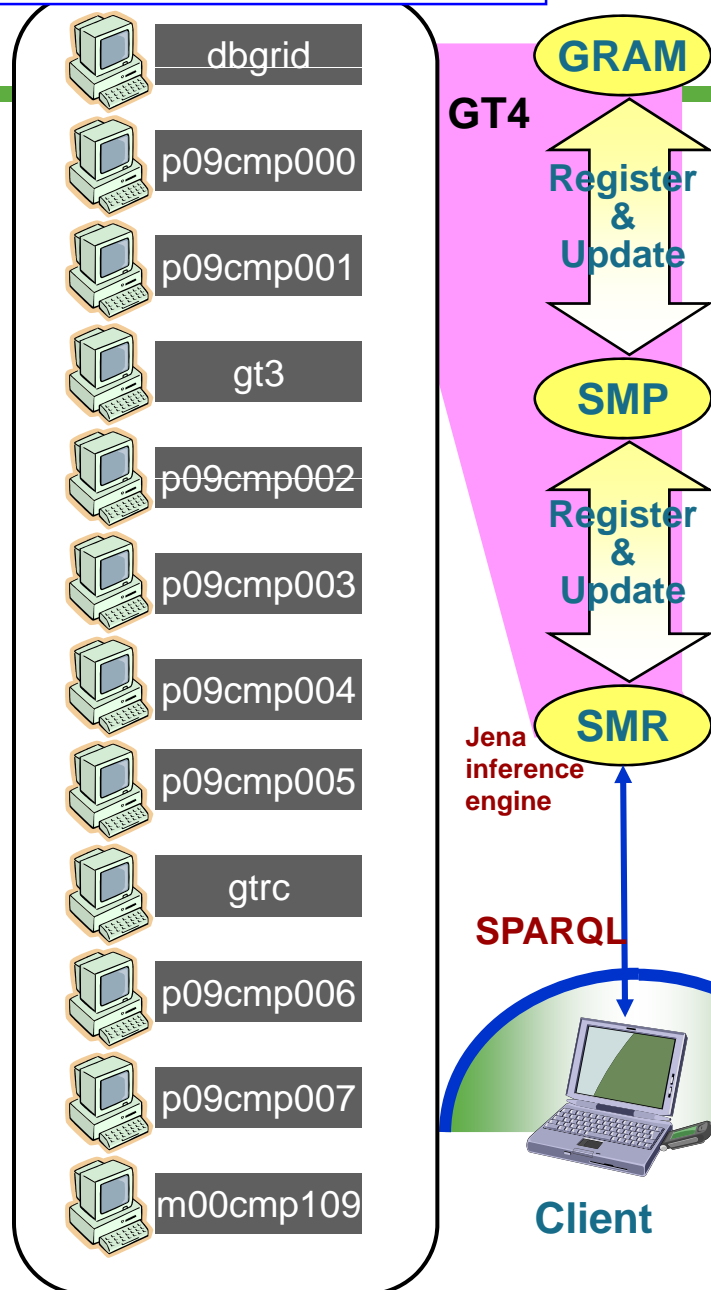


Appendix



Data Grid Cluster (Ganglia & TORQUE)

Monitoring Client



Issues

WS-DAI-RDF(S) Querying

Issues(1)

Harmonize with Ontology Spec



1. Terminology

We should unify the terminology list to read both specs.

Initial “Glossary of Terms” is just given by UPM.

This covers most of the terms used for the specs.

Should be put on the forge after adding some terms.(will be discussed later)

2. Correspondence of Terms between 2 specs.

Most of the terms used within 2 specs seems to have no overlap since these specs cover separate point of views.

(Ont) Repository Resource ?=? (Query) Collection = Set of Graphs with URI

3. Single Namespace?

4. Functional Overlap?

Issues(2) (these are the same with OGF19)

Out of Scope of the current Querying Spec.



1. No Update Language

- No standard Update language for RDF now

2. No Schema Support (only Graphs and Triples)

- Will be supported by Ontology spec.

3. No Reasoning Functions

- Out of the scope of the language

Issues (3) (these are the same with OGF19)

Inside the Query Spec.



1. Two types of interface for one itemsSet is defined
 - Depends on the Result format
 - RDF/XML: TriplesSet
 - SPARQL/XML: ResultsSet
2. External resource flag is introduced.
(please see the document if you interested)
 - Is this really useful?
3. Graph operations are introduced.
 - Only for graphs, Not for triples
 - Includes Update functions

GraphName->GraphNameURI

(which is different from OGF19 version)



Most Important Issue in this Spec

Current: GraphName(string type)

Graph Identity = EPR with GraphName

W3C Standards:

Graph Identity=URI

Change: (GraphName => GraphNameURI)

Graph Identity = EPR with GraphNameURI

Note:

EPR must cope with URI

Uniqueness of the GraphNameURI must be supported.

(These will be the responsibility of the implementation)

String name alias is not defined. (currently)

Issues(4) W3C standards

(They are the same with OGF19)



We must support these standards as possible

- SPARQL
 - No direct relationship at now.
- SPARQL result set XML
 - WS-DAI supports this as a standard data format.
- SPARQL protocol for RDF
 - To be upward compatible
 - Might cause some mismatch
 - Errors, Return format,,,,

Other Issue : Enlarge the Community



- Need to have more authors/contributors.
 - Announced DAIS-WG and Semantic Grid RG
 - Current response is not so good.
 - Still need to advertise this activity.
 - To? How?

Appeal to wider audience

- After the release of the Next version of the Document
- W3C DAWG

