SESSION 1: UPDATE TO THE COMMUNITY (WEDNESDAY 5TH MARCH, 12:00-13:30)

1. Introduction. Review of Activities to Date.

2. Overview of Grid Database Services Proposal.

3. Overview of OGSA-DAI software as reference implementation of Grid Database Services Proposal.

4. Summary/Plans.

1. Dave Pearson reviewed activities to date.

The group is a Working Group and so cannot live beyond 2 years. The main goal is to put forward standards recommendations. The group arose out of a BOF in GGF4, February 2002. For this meeting there is a finalised requirements document, an updated GDS specification, and an updated document on data services.

 OGSA-DAI is a project to develop a reference implementation of the standards definitions.

- Phase 1 Feb-Sept 2002 complete.
- Phase 2 Oct 2002-July 2003.

2. Greg Riccardi. Overview of the Grid Data Services Proposal.

A Draft Specification was produced for GGF6. The draft specification produced for GGF7 is a significant update.

The architecture is the standard OGSA architecture, with registries that return factories, and factories that create a GDS.

The GDS adds two new port types to the standard OGSI services: a GDS port and a GDT (Grid Data Transport) port.

The GDS port provides service data elements that describe both the data resource, and the service.

A set of typical Client-Server Interaction Patterns were presented, along with examples.

Question (MPA): can one query that stores the result in the service be lodged with the service, and then called repeatedly, with each result separately identifiable?

Answer (GR): the current specification does not support that.

Question (Bill Allcock): how long can results be stored at the service.

Answer (GR): that is a matter for the GDS. It may require negotiation between the client and the service.

Question (Guy Rixon): Is it sensible to solve the problem by using one instance of the GDS for each query, so that the lifetime management of the service can be used to control how long the results of a query are stored?

Answer (GR): that could be a solution.

Comment (MPA): There is a challenge in mapping between OGSA security and the security policies of a legacy databases to which a GDS is connected.

GR: The current issue of the specification does not cover security. It will be extended to encompass to cover security.

Comment (??): Changes (rather than just extensions) to the existing specification may be required if the data returned from queries needs to be encrypted.

Question (??): Is there any provision for stored procedures?

Yes, the SQL query statement can handle a procedure call, but in the future stored procedures may be directly supported

Question (Guy Rixon): Is the implementation affected by changes in the OGSI registry implementation?

Answer (Neil Chu Hong): Yes, but only a limited amount.

3. James Magowan. Overview of OGSA-DAI software as reference implementation of Grid Database Services Proposal.

Discussion of what is supported in the current release of OGSA-DAI, and future plans.

4. Dave Pearson, Summary:

It is hoped that the scope will be extended to include files.

Outstanding issues are: data movement, large data sets and security.

Question (Inderpal Narang): Is the resource negotiation required by DQP key for the next phase of OGSA-DAI?

Answer: Currently we do not have solutions to extracting the schema and performance estimates from a database.

1. Discussion of the updated Draft Grid Database Services Specification.

2. Discussion of GridDataTransport portType from specification, with a view to understanding the extent to which this overlaps with transport requirements for other areas.

3. Summary/Plans.

1. Susan Malaika: A Review of the Grid Data Service Specification

Dave Snelling volunteered to work with the specification team to ensure that it is OGSI compliant.

It is necessary to review the spec further in various areas:

- Use Patterns (the Primer)
- Document Structure (Factory and Perform)
- RPC Structure
- Activities
- Integration with Delivery
- Metadata and Service Data

Inderpal Narang noted that control had been removed from the latest draft of the standard. Susan said that this was because of a lack of time.

There is a need to deal with transactional behaviour. Norman Paton suggested that WS-Transactions may be of useful for overall control.

Question (Guy Rixon): Could there be information on the semantics of activities?

Response (Norman Paton) Three types of description are needed: syntax (in specification); description of what syntax allows you to do (in the primer); semantics (in specification).

Guy Rixon volunteered to provide a description of what activities his project required.

Comment (Inderpal Narang): The spec should be reviewed with a workflow expert. Frank Leymann (IBM) was suggested.

Question (James Magowan): The spec should describe how activities can be added, without specifying the activities themselves. The Grid services spec could be used as a model for this.

Question (Inderpal Narang): It is compulsory for databases to provide metadata and service data?

Answer (NP): no the spec says that it should be provided, so databases can choose what to expose.

Comment (MPA): we need an activity to develop metadata to describe databases and their contents, e.g. there is a standard for archives Z3950.

Inderpal Narang from IBM volunteered Vijayshanker Raman, who is active on OGSI security, to provide security advice to the specification team.

Comment(MPA): Suggested there should be a focus on the database aspects of security, rather than look right across the broad area of security.

Primer:

Greg Riccardi would like examples from the grid community that could be included in the primer. The slides will be published tomorrow.

Specification Issues:

- Transformation activities need to be added into the spec

2. Discussion of GridDataTransport portType from specification, with a view to understanding the extent to which this overlaps with transport requirements for other areas.

Malcolm Atkinson gave an overview of issues with data movement.

Inderpal Narang talked about composing the GDS with the Grid Movement & Replication (GMR)

GMR is work in progress within IBM.

Comment (Bill Allcock): A lot of what Malcolm talked about was not transport.

MPA replied that it is information transport rather than data transport.

Question (Guy Rixon): What happens when data is pushed from one service to another - how does the target decide what to call the data?

Answer (Greg Riccardi): The data travels with an identifier.

Question (?? from Yotta Yotta): What happens if data is being updated as it is being transferred?

Answer (Inderpal Narang): Databases provide facilities to maintain coherence: they use the log to identify changes that need to be propagated.

Question (?? from Yotta Yotta): In the storage world, hot copies are required.

Answer (Norman Paton): in the database world read consistency is required, and so a consistent snapshot of the data is required by queries.

Question (???): Is it the job of this group to define a replica?

Answer (all): No. There is another group discussing this. However, this group should ensure that no design decisions are made that prevent higher-level replica management services from operating over GDS.

Question (Dave Berry): Is there any intention to specify the semantics of the terms used in the specification?

Answer (Norman Paton): The spec will not include a formal specification, but it will become more precise than it is at the moment.

Comment (MPA): The target of Version 1.0 of the specification ready for GGF8 may be too ambitious. We may have to defer until GGF9 if there is not time to fully discuss it and resolve issues before GGF8.

Minutes: Paul Watson (5th March 2003).