UK e-Science
OGSA-DAI

Malcolm Atkinson
Director of National e-Science Centre
www.nesc.ac.uk

Neil Chue Hong
Technical Manager OGSA-DAI
www.epcc.ed.ac.uk

November 2002
SuperComputing 2002
Baltimore
OGSA-DAI

Data Access and Integration for the New Grid

Uniform Service Interfaces for Accessing Multiple Data Sources within the Open Grid Services Architecture.

A UK e-Science Programme’s Contribution to GT3
A Multi-Centre, Multi-Company Collaboration

Funded by the UK e-Science Grid Core Programme

to Develop Data Access and Integration Services for OGSA

- Access to XML Databases -
- Access to Relational Databases -
- Distributed Query Processing -
- XML Schema Support for e-Science -
OGSA-DAI Partners

EPCC & NeSC
IBM UK
IBM USA
Manchester e-SC
Newcastle e-SC
Oracle

$5 million, 18 months, started February 2002
Building on OGSA & OGSI

- Powerful rallying cry
- Well organised
  - OGSI, OGSA, ...
- Responsive to DAI
- Substantial industrial investment
- Multi-national intensely active working groups
- IBM announcement of contribution of its OGSI code to public source
- Technical Previews out
- Alpha 15th January ‘03

- Major engineering effort
- Complex & large code base
- More demanding of software environment
  - Web services
  - Messaging
- Depends on planned extensions to standards
- First APIs for Java

Gets our vote
GGF DAIS WG

- **Chairs**
  - Norman Paton (Manchester Uni.)
  - Leanne Guy (CERN) ... dropped out
  - Dave Pearson (Oracle UK)

- **Activity**
  - BoF GGF4 Toronto
  - WG Meeting GGF5 Edinburgh
  - Papers for GGF6
  - Workshops & Mail lists

- **Goals**
  - Agree Standards for Database Access & Integration
  - Freely available reference implementations
    - OGSA-DAI one source & focus for discussions

- **Data Provenance and Derivation Workshop**
  - Significant UK contribution

http://www.cs.man.ac.uk/grid-db/
OGSA-DAI Stake Holders

Data Intensive Application Scientists

Keep all the stake holder groups satisfied
DAI Key Services

- **GridDataService** (GDS): Access to data & DB operations
- **GridDataServiceFactory** (GDSF): Makes GDS & GDSF
- **GridDataServiceRegistry** (GDSR): Discovery of GDS(F) & Data
- **GridDataTranslationService** (GDTS): Translates or Transforms Data
- **GridDataTransportDepot** (GDTD): Data transport with persistence

Integrated Structured Data Transport
Relation & XML models supported
Role-based Authorisation
Binary structured files (later)
1a. Request to Registry for sources of data about “x”

1b. Registry responds with Factory handle

2a. Request to Factory for access to database

2b. Factory creates GridDataService to manage access

2c. Factory returns handle of GDS to client

3a. Client queries GDS with XPath, SQL, etc

3b. GDS interacts with database

3c. Results of query returned to client as XML
Composing Components

OGSA-DAI Component

Data Transport

OGSA-DAI Component

Data Transport

OGSA-DAI Component

Data Transport

OGSA-DAI Component

Data Transport
Distributed Query

DQP : Distributed Query Processor
GDT : Grid Data Transport
T : Translation
Q : Query
GDTV : Grid Data Transport Vehicle
F : Factory
QPM : Query Progress Monitor
PNM : Progress Notification Message
AM : Application Metadata
CRM : Computational Resource Metadata
NS : Notification Sink
Interface transparency:

one GDS supports multiple database types
OGSA-DAI Time Line

Phase 1 Starts
Feb ’02
May ’02
Jul ’02
Sep ’02

Phase 2 Starts
Dec ’02
Feb ’03
May ’03
Sep ’03

WS + GSI UK support (> 100 downloads)
XML + OGSA Prototypes for Early Adopters

Design Documents & Demos for DAIS WG @ GGF5

XML + OGSA Prototype Available
RDB + GT2 / OGSA Prototypes Available
GGF6 WG Papers & Prototypes

Ship Release 1 for GT3 Integration
Demo & Release 1.5 @ GGF7

Release 2
today
End of Presentation

Questions Please
Software Availability

Available now

- Phase 1 prototype of GDS, GDSF & GDSR for XML
  Java implementations for the axis/tomcat platform and the Xindice database
- Globus-2 Relational database support
- BinX Schema v0.2
  [www.epcc.ed.ac.uk/gridserve/WP5](http://www.epcc.ed.ac.uk/gridserve/WP5)
  An XML Schema for describing the structure of binary datafiles – the power of XML for terabyte files

Software Q1 2003

- Reference implementation 1
- Access & Update
  - XML databases
  - Relational databases
- To be released as Basic Services in Globus Toolkit 3

[umbriel.dcs.gla.ac.uk/NeSC/general/projects/OGSA_DAI/products](http://umbriel.dcs.gla.ac.uk/NeSC/general/projects/OGSA_DAI/products)
<table>
<thead>
<tr>
<th>Class</th>
<th>GridService</th>
<th>Registry</th>
<th>NotificationConsumer</th>
<th>NotificationProducer</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDS</td>
<td>Mandatory</td>
<td>Optional</td>
<td>Normal</td>
<td></td>
</tr>
<tr>
<td>GDSF</td>
<td>Mandatory</td>
<td>Optional</td>
<td>Normal</td>
<td></td>
</tr>
<tr>
<td>GDSR</td>
<td>Mandatory</td>
<td>Mandatory</td>
<td>Normal</td>
<td></td>
</tr>
<tr>
<td>GDTS</td>
<td>Mandatory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDTD</td>
<td>Mandatory</td>
<td>Optional</td>
<td>Normal</td>
<td></td>
</tr>
</tbody>
</table>
# DAI portType Usage

<table>
<thead>
<tr>
<th>Class</th>
<th>GridDataService</th>
<th>DataTransport</th>
<th>Factory</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDS</td>
<td>Mandatory</td>
<td>Normal</td>
<td></td>
</tr>
<tr>
<td>GDSF</td>
<td>Optional</td>
<td>Normal</td>
<td>Mandatory</td>
</tr>
<tr>
<td>GDSR</td>
<td>Optional</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GDTS</td>
<td>Optional</td>
<td>Mandatory</td>
<td></td>
</tr>
<tr>
<td>GDTD</td>
<td>Optional</td>
<td>Mandatory</td>
<td></td>
</tr>
</tbody>
</table>
Advanced Components

Client → GDS:PerformScript → GDS → GDT → DB → Translation → Translation → Consumer
Composed Components