Outline of Talk

- Project background and motivation
- Status of current release
- Releases and Usage Statistics
- Roadmap for future releases
Entering an age of data

- Data Explosion
  - CERN: LHC will generate 1GB/s = 10PB/y
  - VLBA (NRAO) generates 1GB/s today
  - Pixar generate 100 TB/Movie
- Storage getting cheaper

Data stored in many different ways

- Data resources
  - Relational databases
  - XML databases
  - Flat files

Need ways to facilitate

- Data discovery
- Data access
- Data integration

Empower e-Business and e-Science

- The Grid is a vehicle for achieving this

Major challenge: rate of growth in number & complexity of data sources & rate at which they change ⇒ old “cottage industry” solutions don’t scale
First steps towards a generic framework for integrating data access and computation
Using the grid to take specific classes of computation nearer to the data
Kit of parts for building tailored access and integration applications
Investigations to inform DAIS-WG
One reference implementation for DAIS
Releases publicly available NOW
Funded by the Grid Core Programme

**OGSA-DAI**
- £3 million, 18 months, from Feb 2002
  - Three major releases, three interim releases

**DAIT (DAI-Two)**
- Keep the OGSA-DAI brand name
  - £1.5 million, 24 months, from Oct 2003
  - Four major releases

**GGF DAIS WG**
- Strong involvement.
- Standardise the interfaces
- OGSA-DAI to be a reference implementation

Project Partners

http://www.ogsadai.org.uk
Project Membership

Research Team

Principal Investigators

Programme Management Board Chair

Technical Review Board Chair

Project Manager

EPCC Team

IBM Development Team

IBM Dissemination Team

http://www.ogsadai.org.uk
**Data Intensive Applications for Science X**

**Simulation, Analysis & Integration Technology for Science X**

**Generic Virtual Data Access and Integration Layer**

- **Job Submission**
- **Registry**
- **Data Transport**
- **Brokering**
- **Banking**
- **Resource Usage**
- **Workflow**
- **Authorisation**
- **Transformation**

**Structured Data**

**Structured Data Integration**

**Structured Data Access**

**OGSA-DAI**

**Grid or Web Service Infrastructure**

- **Compute, Data & Storage Resources**

**Distributed Virtual Integration Architecture**
Request to Registry for sources of data about “x”

Registry responds with Factory handle

Request to Factory for access to database

Factory returns handle of GDS to client

Factory creates GridDataService

Client queries GDS with SQL, XPath, XQuery etc

Query results returned in XML

OR delivered to consumer as XML

GDS interacts with database

Database (Xindice, MySQL, Oracle, DB2)
Multiple tasks / request

Data Set

Ident
Type
Value

1

2

Ident
Type
Value

Ident
Type
Value

Ident
Type
Value

Ident
Type
Value

Ident
Type
Value

Ident
Type
Value
Current status

- OGSA-DAI R4 available since April 2004
- OGSI based – built on top of GT 3.2
- Supports relational, xml and some files
  - MySQL, Oracle, DB2, SQL Server, Postgres, XIndice, CSV
- Supports various delivery options
  - SOAP, FTP, GridFTP, HTTP, files, email, inter-service
- 2746 downloads, 792 registered users (Aug 04) from all around the world
- If you need to build high level data services in your project – we strongly urge you to build on OGSA-DAI

http://www.ogsadai.org.uk
R4 April 2004

- Provides Data Access components, an extensible framework for building applications and some integration components
- Built on top of Globus Toolkit 3.2
- Supports relational, xml and some files
  - MySQL, Oracle, DB2, SQL Server, Postgres, XIndice, CSV
- Supports various delivery options
  - SOAP, FTP, GridFTP, HTTP, files, email, inter-service
- Supports various transforms
  - XSLT, ZIP, GZip
- Supports message level security using X509 certificates
- Client Toolkit library for application developers
- GUI data browser (contributed by FirstDIG project)
- Separate Distributed Query Processing components
- Comprehensive documentation and tutorials in XHTML format
Downloads by country

792 registered users @ 23/8/04
Downloads by Release

2746 downloads (~4.7 downloads a day)
Projects using OGSA-DAI at UK 2004 AHM

- Astrogrid
- ConvertGrid
- eDiamond
- EdSkyQueryG
- FirstDIG
- GEDDM
- GeneGrid
- INWA
- myGrid
- ODD-Genes

All these projects have demos at All Hands
Some more projects

“Expediting the cancer research communities' access to key bioinformatics platforms by deploying an integrating biomedical informatics infrastructure”

- Chosen to use OGSA-DAI and Project Mobius to create data infrastructure

OGSA-DAI being deployed on all GEON nodes
Future plans

OGSA-DAI is working on:
- WS-I/WS-RF interfaces
- Data integration tools and applications
- Workflow and mobile code

Basic WS-I version will be released with OMII middleware in October 2004
- OGSA-DAI R5 (GT 3.2.1 based) will be available October 2004 also

OGSA-DAI R6 available April 2005
- Full support for WS-RF
- Data Integration applications supporting identified scenarios
- OGSA-DQP as an integrated part of release
- Fully compliant JDBC Driver for OGSA-DAI
- Support for WS-Security implementations … and much more

http://www.ogsadai.org.uk
New roadmap document published on OGSA-DAI website
  - http://www.ogsadai.org.uk/docs/OtherDocs/OGSA-DAIRoadmapV2.0.pdf

User feedback required to drive this document
  - Suggest requirements
  - Let us know your priorities

User Group meetings
  - Next one at GGF12 (Monday 20th September, 5pm)
R5 October 2004 (Interfaces) – OGSI

- Built on Globus Toolkit 3.2.1
- Re-engineered interface-independent core OGSA-DAI functionality.
- Improved dependability and security integration.
- New file data resources representing flat files queried using full text searches (e.g. EMBL format).
- Installation and Configuration Wizard, including “all-in-one installer”
- Improved Data Browser which allows XPath querying.
- Set of standard benchmarks.
- JSP Quick View interface.
- Support for other databases (e.g. Access, Exist, HSQL).
WS-I Technical Preview

- A limited functionality evaluation version
  - An OGSA-DAI “Data Service” combining the metadata, configuration and perform document capabilities of the OGSI-based GDSF and GDS services.
  - Access to service metadata provided by a partial implementation of the WS-ResourceProperties specification.
  - Example clients are provided for testing and coding reference.

- Caveats/Issues:
  - No registry component, no support for third party delivery.
  - Security may be available (based on OMII WS-Security plug-in for Axis).
  - Document schema and interfaces WILL change.
  - The WSDL is based on the OGSI-based WSDL from OGSA-DAI
  - Will not be supported to same level as main release.

- Released with OMII middleware distribution in October
R6 April 2005 (Integration) – WS-RF

- Data Integration applications supporting identified scenarios
- OGSA-DQP as an integrated part of release
- Fully compliant JDBC Driver for OGSA-DAI
- Support for WS-Security implementations
- Support for stored procedures on all supported databases
- Improved support for different database specific SQL types
- SQL translation between vendor dialects for subset of queries
- Support for XQuery data resources
- We expect to comply with a version of the emerging DAIS specification at this release.
Future releases

- Produce a reference implementation of the DAIS Specification
- Integrate with other eScience components
  - Workflow is vital
- Start to provide tools to manage distributed data resources
- Collaborate with other groups to produce additional functionality
Data Access – done

Data Integration – in progress

OGSI → WS-I → WS-RF

OGSA-DAI provides a framework which reduces application development time

We need to know what you’re doing
  – Comment on the Roadmap
  – Participate in the User Group

OGSA-DAI is out there NOW
  – Use it, and let us know what you’ve done
Further information

- The OGSA-DAI Project Site:  
  - http://www.ogsadai.org.uk

- The DAIS-WG site:  
  - http://cs.man.ac.uk/grid-db

- OGSA-DAI Users Mailing list  
  - users@ogsadai.org.uk  
  - General discussion on grid DAI matters

- Formal support for OGSA-DAI releases  
  - http://www.ogsadai.org.uk/support  
  - support@ogsadai.org.uk

- OGSA-DAI training courses