



Enabling Grids for
E-science in Europe

www.eu-egee.org

GDA meeting at CERN, 10 May 2004

R-GMA status and plans

Steve Fisher / RAL



Contents

- Who we are
- Reminder of R-GMA
- What we have done so far
- What we plan for the first release of the EGEE middleware



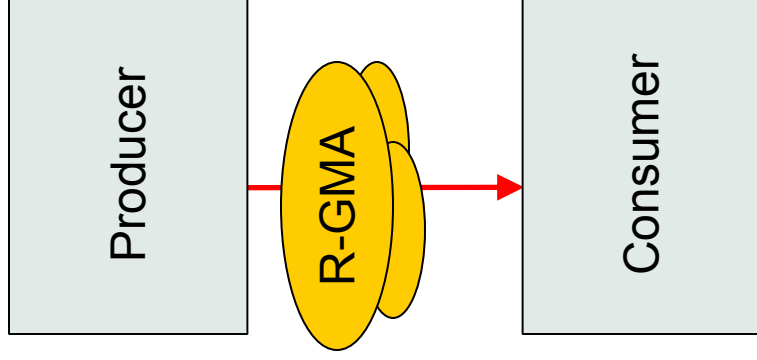
The mission and the people

- **Mission**
 - To meet all information and monitoring needs
 - Expect to base much of our re-engineering work on R-GMA
 - Inter-work with other information systems and displays
 - We are not responsible for the contents of the schema but will coordinate as necessary
- **The People**
 - 8 FTEs at RAL and 1 more to recruit
 - 1 FTE at QMUL (unfunded)
- **We can be e-mailed at:**
 - jra1-uk@physics.gla.ac.uk

Unified approach to information and monitoring

- Provides information on both:
 - The Grid itself
 - Mainly for the middleware packages
 - The user may query it to understand the status of the Grid
 - Grid applications
 - For users

R-GMA



- A relational implementation of GMA (from GGF)
 - Powerful data model and query language
 - All data modelled as tables
 - SQL can express most queries in one expression
- Applied to both information and monitoring
 - All tuples have time stamp
- Creates impression that you have one RDBMS per VO

Relational Data Model

- **Not** a general distributed RDBMS system, but a way to use the relational model in a distributed environment where **global consistency is not important**.
- **Producers** announce: SQL “CREATE TABLE”
- **publish:** SQL “INSERT”
- **Consumers** collect: SQL “SELECT”
- Some Producers (and the Registry and Schema) make use of RDBMS as appropriate – but what is central is the **relational model**.

Producers

- StreamProducer – Supports Continuous Queries
 - In memory data structure
 - Can define minimum retention period
- DataBaseProducer – Supports History Queries
 - Information is retained – has a clean up strategy
 - Supports joins
- LatestProducer – Supports Latest Queries
 - Just holds the latest information for any “primaryish” key
 - Supports joins
- CanonicalProducer – Supports anything
 - Offers “anything” as relations
 - User has to write code to handle SQL etc.

**This is all
much cleaner
in the new
EGEE design**

Archiver (SecondaryProducer)

- It is a combined Consumer-Producer
 - Follows the GMA concept but packaged for ease of use
- You just have to tell it what to collect and it does so on your behalf

R-GMA Tools

- R-GMA Browser
 - Application dynamically generating web pages
 - Supports pre-defined and user-defined queries
- R-GMA CLI (edg-rgma)
 - Command Line Interface (similar to MySQL)
 - Supports single query and interactive modes
 - Can perform simple operations with Consumers, Producers and Archivers

LCG Installation of R-GMA

- We have produced RPMs and instructions for adding R-GMA to LCG2
- This has been done successfully at a number of sites in UK and elsewhere.

JRA1: Milestones and Deliverables for the first year

Month	Deliverables & Milestones	Item	Lead Partner
M03	MJRA1.1	Tools for middleware engineering and integration deployed	CERN
M03	DJRA1.1	(Document) Architecture and Planning (Release 1)	CERN
M03	MJRA1.2	Software cluster development and testing infrastructure available	CERN
M05	MJRA1.3	Integration and testing infrastructure in place including test plans (Release 1)	CERN
M05	DJRA1.2	(Document) Design of grid services (Release 1)	CERN
M09	MJRA1.4	Software for the Release Candidate 1	CERN
M10	MJRA1.5	Integrated Release Candidate 1 enters testing and validation period (Release 1)	CERN
M12	DJRA1.3	(Software) Software and associated documentation (Release 1)	CERN

The problem

- We need to re-engineer existing code, and introduce new functionality while at the same time providing continuously improved frequent releases and all without breaking backwards compatibility.

Activities so far

- New API implemented (partially) as wrapper round the old one for the EGEE MW prototype
 - We would value feedback on the new API
- Registry replication completed and tested on our cluster testbed
- The new API includes authorisation
 - Not complete yet – but a good start
 - Nothing behind it yet
- Web service prototypes written for all components
- Extensive UML diagrams for new design
 - Once design agreed we will refactor code towards it
- Outline plan agreed in for Schema replication

Overview of tasks for the year

- Introduce multi-VO support
- Complete the authorisation design (API and internally)
- Complete the schema replication
- Enhance the mediator
- Move to web services
- Clean up the SQL parsing so that it corresponds precisely to one of the SQL92 levels
- Watch developments elsewhere and seek to inter-work
- Work with others to understand relationship between new services and I&M.

May and June

- **May 31**
 - Detailed specification of the software. This is expected to be around 30 pages and will attempt to tie down all the undocumented uncertainties in the behaviour of the system.
- **June 15**
 - Code will have been moved to the new repository, suitably restructured for the new SCM and build system. We will have had to learn the build and SCM system and possibly wait for bugs to be fixed in this system.

- **July 31**
 - The Java wrapper will be finalized (in terms of the API that is) along with wrappers in C++, C and Python.
 - API will exploit the features of each language rather than going for a lowest common denominator approach. This will take a bit longer but will result in a better product.
 - After this we will endeavour not to change the API further, though we may make additions if these prove desirable.
- **July 31**
 - A complete design of the system including schema replication and the support of multiple VO's so that we can migrate/refactor to this agreed design.
 - This design will be updated a little as we go along and will represent our goal.

August to December

- **Oct 31**
 - The whole system will be ported to Windows
- **Oct 31**
 - We will make it very easy to build and distribute the system in source form such that it can be readily built and installed on new platforms.
- **Dec 20**
 - Multiple releases will be made with some refactoring, culminating in a release with multiple VO support and Schema replication and where the servlets are replaced by Web Services. Some improvements will be made to the mediator and to the SQL parser.

Over the year

- During the year:
 - Authorization code will be developed but it is not yet clear how much of it will be ready for RC1.
 - We will also look at interfacing to various displays and other information and monitoring systems
 - by working with other clusters as appropriate understand the R-GMA role in Accounting, Job provenance and Logging & bookkeeping.

And finally

- Note that after July 31st there will be no external changes to the software so we will not be under pressure to release code before it is ready.
- We may release less improvements than planned by Dec 20th **but we will deliver a working system.**