

**ICEAGE**




**eGee** Enabling Grids for E-scienceE

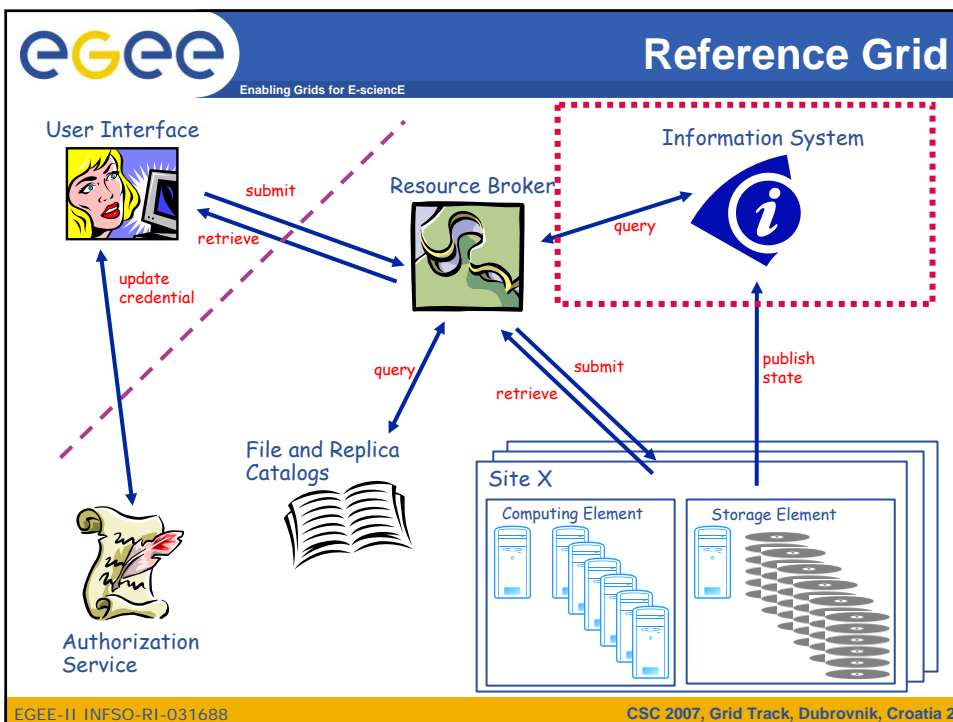
# Information Systems

*Heinz Stockinger*  
*Swiss Institute of Bioinformatics*

[www.eu-egee.org](http://www.eu-egee.org)

EGEE-II INFSO-RI-031688 EGEE and gLite are registered trademarks



## Information Systems in General

### One Specific Implementation

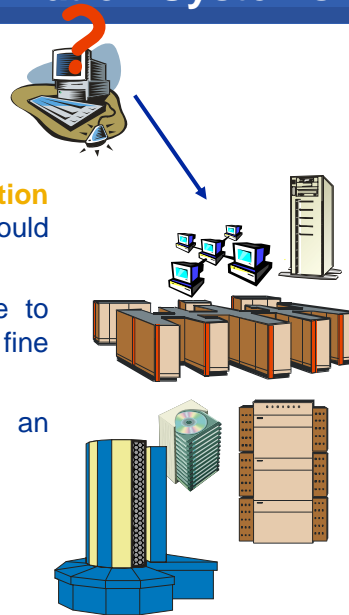
R-GMA

### Schema

## Information Systems

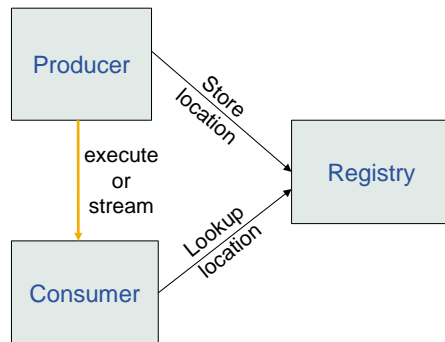
- **Some Requirements**

- The system must be able to cope with nodes in a **distributed environment**
- Dynamic **addition and deletion information providers** (producers) should be supported
- It must have a **security system** able to address the access to information at a fine level of granularity
- It must be able to work well on an unreliable **Wide Area Network (WAN)**
- Scalability, performance, reliability



- **Provides information on both:**
  - The **Grid** itself (i.e. the basic resources and services)
    - Mainly for the middleware services
    - The user may query it to understand the status of the Grid
  - Grid **applications** (for users)
- **Schema required**
  - To describe all resources in the Grid
  - Example: Computing resource and available CPU
- **Query “language” required**
  - Features are similar to a Database Management System
- **Several architectural approaches**
  - No common agreement yet but several standardisation efforts
  - Standards are required to make Grids interoperable
    - Analogy: SQL/ODBC can be used to query database implementations of several different vendors

- **Many Grid projects provide their own implementations:**
  - Globus: MDS (Monitoring and Discovery System)
  - EGEE: bdII/R-GMA (Relational Grid Monitoring Architecture)
  - UNICORE: Incarnation Database (IDB)
  - NAREGI: CIM (industry standard)
  - etc.
- **A service registry is used to discover (Grid) services**
  - Need for standardisation since a **service registry is required by all Grid systems that use service oriented architectures**
  - Also referred to as **service discovery**
- **Grid Monitoring Architecture (GMA)**
  - For information storage and general monitoring of services and status
  - Standardisation in OGF
  - One of the early standard proposals in OGF



- From OGF
- Very simple model
- Does not define:
  - Data model
  - Data transfer mechanism
  - Registry implementation

Defined in OGF Document: GFD-I.7

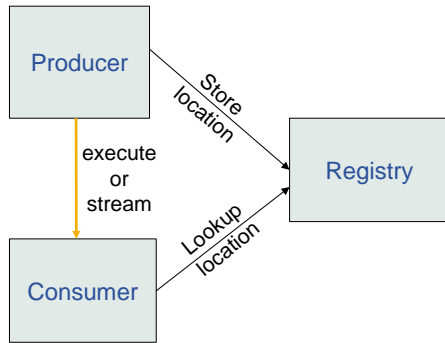
<http://forge.gridforum.org/projects/ggf-editor/document/GFD-I.7>

## Information Systems in General

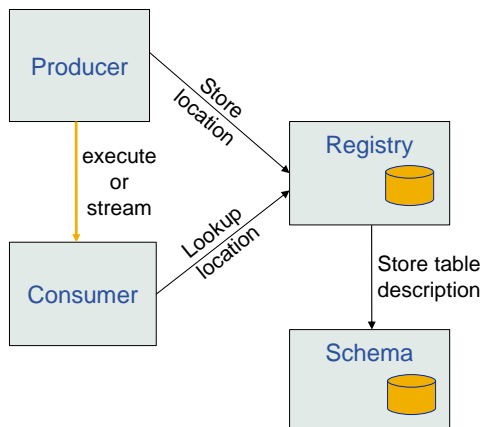
### Specific Implementation

R-GMA

Schema



- Use the GMA from OGF
- A relational implementation
  - 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
- Creates impression that you have one RDBMS per VO



- Registry has two main tables:
  - Producer
  - Consumer
 Only used to discover Producers/Consumers
  - Query goes then directly from consumer to producer
- Schema holds description of tables
  - Column names and types of each table

- **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 “CREATETABLE”  
publish: SQL “INSERT”
- **Consumers** collect: SQL “SELECT”
- Some producers, the Registry and Schema make use of RDBMS as appropriate – but what is central is the relational model.
- All R-GMA tuples are time-stamped
- Uses web service technology

- **R-GMA Browser**
  - Application dynamically generating web pages
  - Supports pre-defined and user-defined queries
- **R-GMA CLI (rgma)**
  - Command Line Interface (similar to MySQL)
  - Supports single query and interactive modes
  - Can perform simple operations with Consumers and Producers
- **Several language bindings in C/C++, Java, Perl etc.**

- \$ rgma

Welcome to the R-GMA virtual database for Virtual Organisations.

=====

Your local R-GMA server is:

`https://lxb1940.cern.ch:8443/R-GMA`

You are connected to the following R-GMA Registry services:

`https://lxb1940.cern.ch:8443/R-GMA/RegistryServlet`

You are connected to the following R-GMA Schema service:

`https://lxb1940.cern.ch:8443/R-GMA/SchemaServlet`

Type "help" for a list of commands.

- rgma> show tables;

Table Name	
GlueCE	Computing Element
GlueCEAccessControlBaseRule	
...	
GlueHostRemoteFileSystem	
...	
GlueSE	Storage Element
GlueSEAccessProtocol	
GlueSEAccessProtocolSupportedSecurity	
...	
Site	Service Discovery
ServiceStatus	
Service	
ServiceData	
VOService	

- `rgma> describe GlueCE;`

Column name	Column type
UniqueID	VARCHAR(128)
Name	VARCHAR(255)
GlueClusterUniqueID	VARCHAR(100)
TotalCPUs	INTEGER
LRMSType	VARCHAR(255)
LRMSVersion	VARCHAR(255)
GRAMVersion	VARCHAR(255)
HostName	VARCHAR(128)
GatekeeperPort	VARCHAR(128)
RunningJobs	INTEGER
WaitingJobs	INTEGER
TotalJobs	INTEGER
Status	VARCHAR(255)
WorstResponseTime	INTEGER
EstimatedResponseTime	INTEGER
FreeCpus	INTEGER
Priority	INTEGER
MaxRunningJobs	INTEGER
MaxTotalJobs	INTEGER
MaxCPUtime	INTEGER
MaxWallClockTime	INTEGER
InformationServiceURL	VARCHAR(128)
MeasurementDate	DATE
MeasurementTime	TIME

24 Rows in set

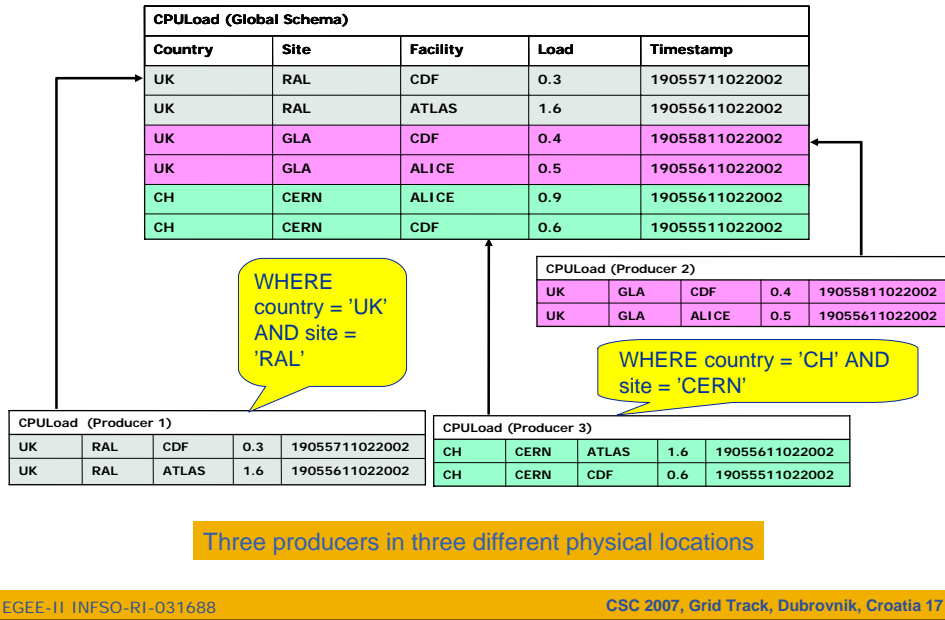
- `rgma> select Name, Site_Name from Service;`

Name	Site_Name
lxb1928.cern.ch_CE_LocalLogger	lxb1940.cern.ch
lxb2090.cern.ch_WMS_LocalLogger	lxb1940.cern.ch
lxb2090.cern.ch_WMS_LogMonitor	lxb1940.cern.ch
lxb2090.cern.ch_WMS_JobController}	lxb1940.cern.ch
lxb1924.cern.ch_CE_Gatekeeper	lxb1940.cern.ch
lxb1924.cern.ch_CE_LocalLogger	lxb1940.cern.ch
lxb1924.cern.ch_TORQUE_maui	lxb1940.cern.ch
lxb1924.cern.ch_TORQUE_pbs_server	lxb1940.cern.ch
lxb1924.cern.ch_CE_Monitor	lxb1940.cern.ch
lxb2090.cern.ch_WMS_ProxyRenewal	lxb1940.cern.ch
lxb2090.cern.ch_Logging_Bookkeeping_Server	lxb1940.cern.ch
lxb2090.cern.ch_WMS_WorkloadManager	lxb1940.cern.ch
lxb1928.cern.ch_TORQUE_maui	lxb1940.cern.ch
lxb1928.cern.ch_TORQUE_pbs_server	lxb1940.cern.ch
lxb1928.cern.ch_CE_Gatekeeper	lxb1940.cern.ch
lxb1928.cern.ch_CE_Monitor	lxb1940.cern.ch
lxb2090.cern.ch_WMS_NetworkServer	lxb1940.cern.ch

17 Rows in set



## Contributions to the "Global" Table (Aggregation)



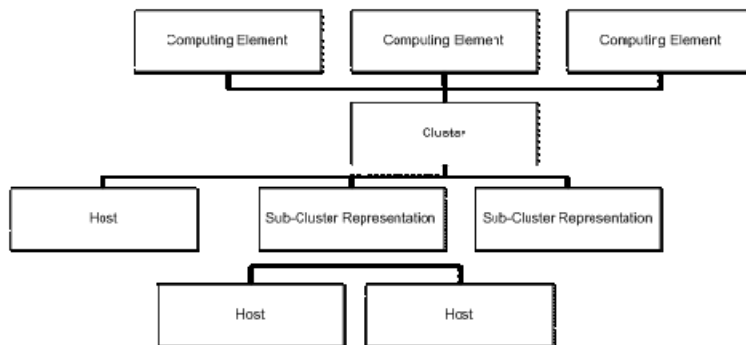
## Information Systems in General

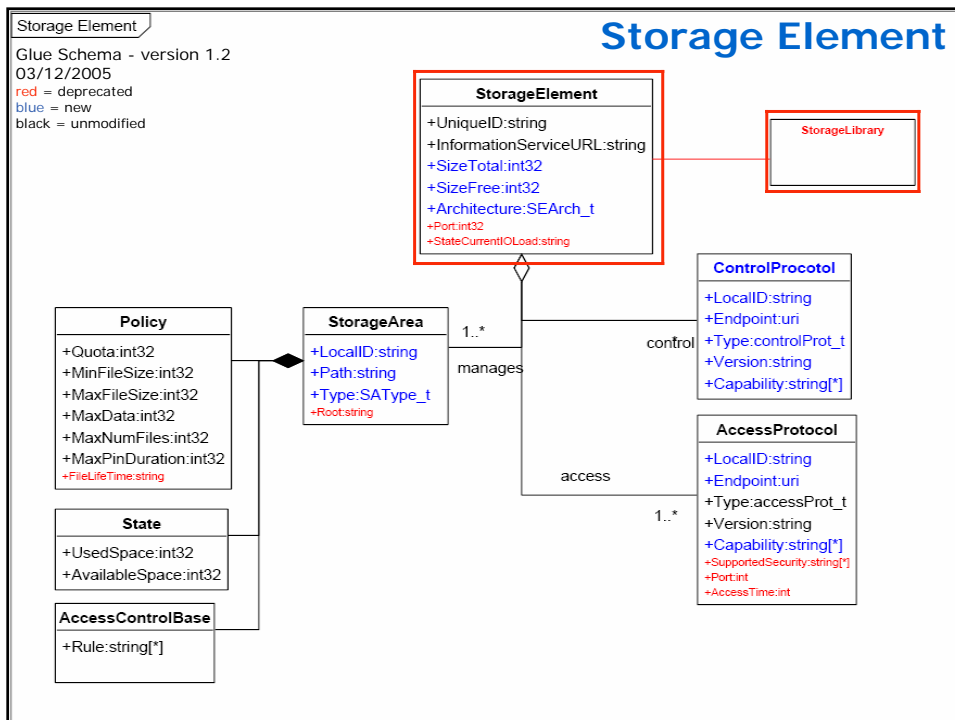
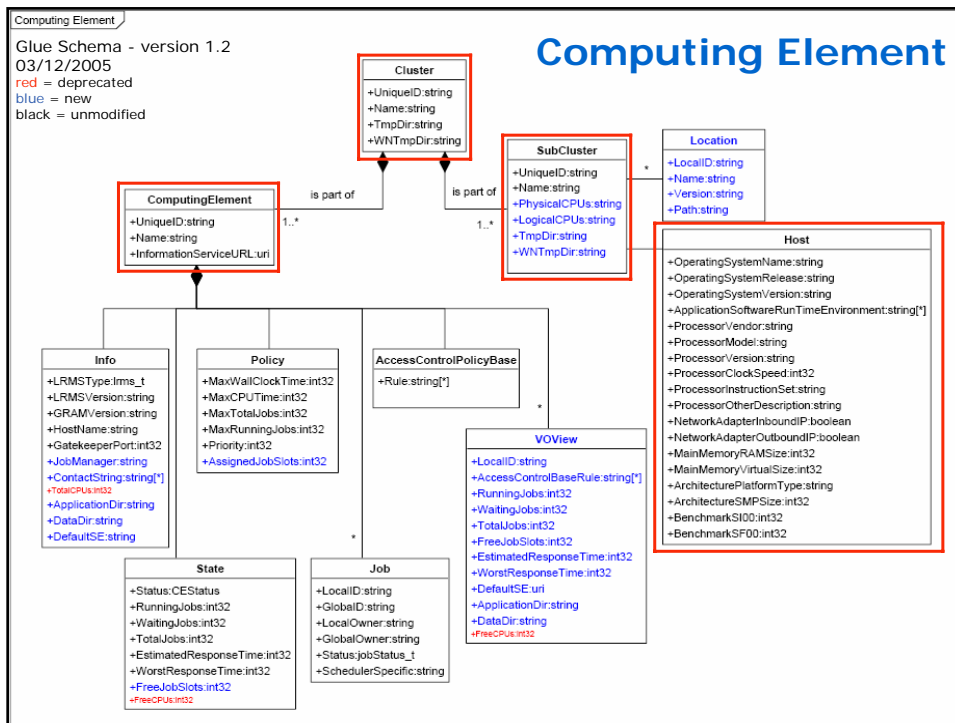
### Specific Implementation

R-GMA

### Schema

- **Developed within High Energy Physics (HEP) community**
  - DataGrid / EGEE
  - DataTAG
  - Globus
- **Currently defines CEs and SEs**
- **Now being internationally standardized via OGF**
- **Entire R-GMA Schema (not only GLUE):**
  - For service discovery and monitoring
  - <http://hepunix.rl.ac.uk/egee/jra1-uk/glite-r1/schema/index.html>





- **R-GMA overview page.**
  - <http://www.r-gma.org/>
- **R-GMA in EGEE**
  - <http://hepunix.rl.ac.uk/egEE/jra1-uk/r-gma/>