Information and Monitoring

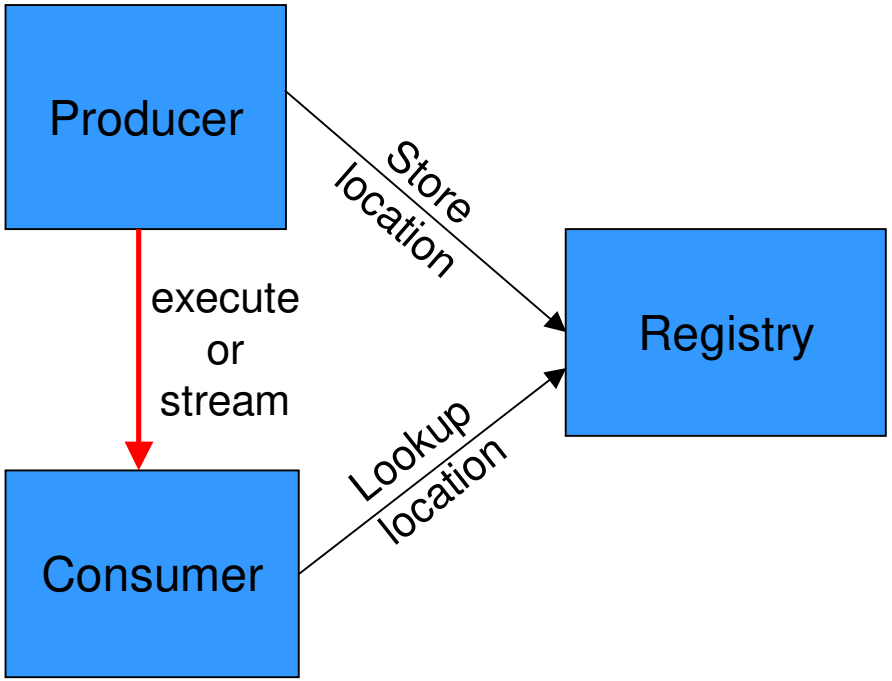
The European DataGrid Project Team

http://www.eu-datagrid.org

Contents

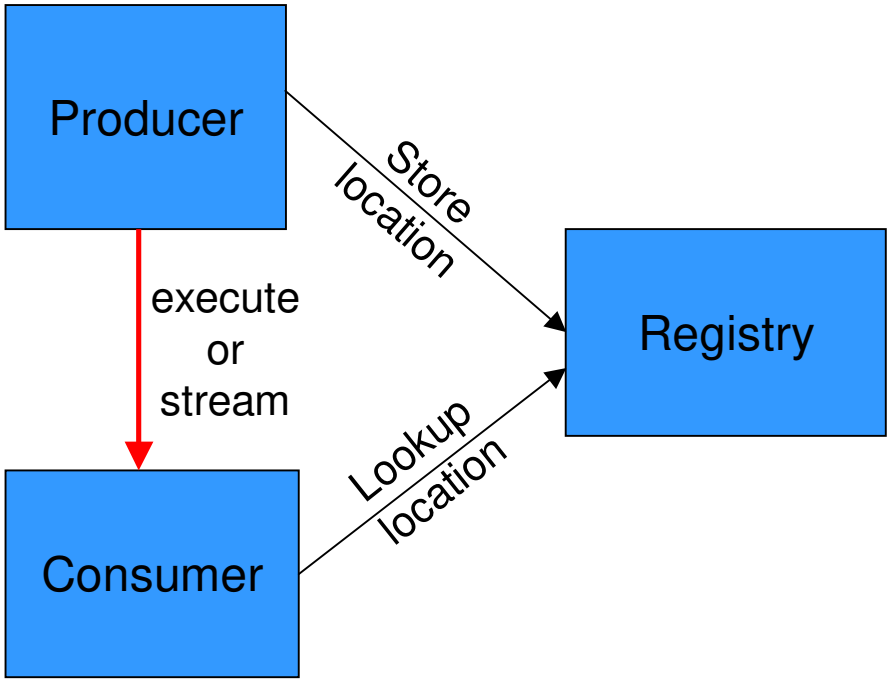
- Scope of Information and Monitoring
- R-GMA
- The schema and the “info-providers”

GMA



- From GGF
- Very simple model
- Does not define:
  - Data model
  - How data are moved from Producer to Consumer
  - What registry looks like

R-GMA



- Use the GMA from GGF
- A relational implementation
  - Powerful data model and query language
- Applied to both information and monitoring
- 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, the Registry and Schema make use of RDBMS as appropriate – but what is central is the relational model.

Producer ➡ Consumer

- Consumer can issue one-off queries
  - Similar to normal database query
- Consumer can also start a continuous query
  - Requests all data published which matches the query
    - Can be seen as an alert mechanism

Producers

- DataBaseProducer – Supports History Queries
  - Information not lost
  - Supports joins
  - Clean up strategy
- StreamProducer – Supports Continuous Queries
  - In memory data structure
  - Can define minimum retention period
- ResilientStreamProducer – Supports Continuous Queries
  - Like the StreamProducer but won't lose data if system crashes
  - So slightly slower
- LatestProducer – Supports Latest Queries
  - Just holds the latest information for any "primaryish" key
  - Supports joins
- CanonicalProducer – Supports anything
  - Offers anything as relations

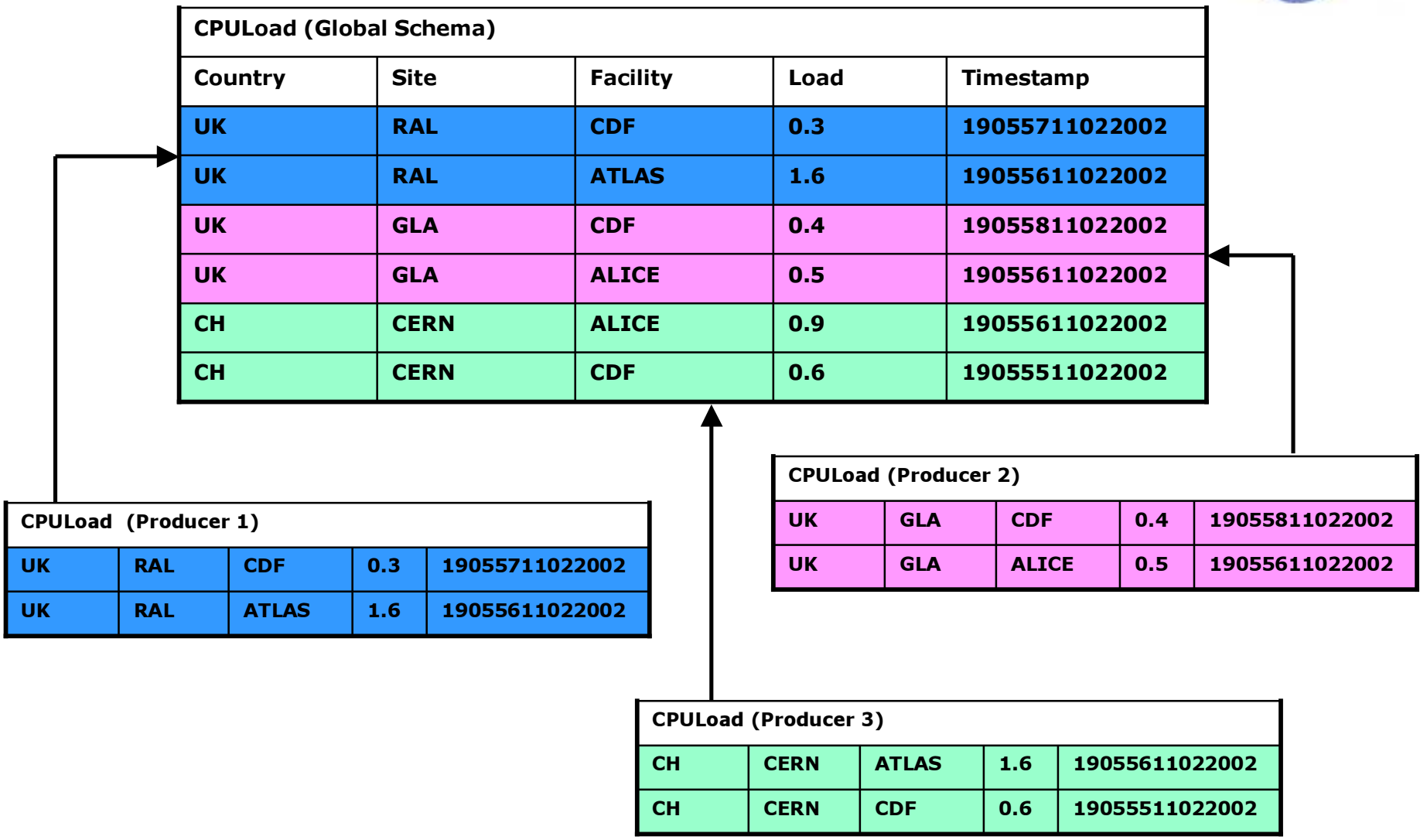
Archiver (Re-publisher)

- It is a combined Consumer-Producer
- You just have to tell it what to collect and it does so on your behalf
- Re-publishes to any kind of "Insertable" (i.e. not to the CanonicalProducer)
  - Can support joins if archiving to a DataBaseProducer or a LatestProducer

Functionality - mediator

- Queries posed against a virtual data base
- The Mediator must:
  - find the right Producers
  - combine information from them
- Hidden component – but vital to R-GMA

Contributions to the "global" table



CPULoad (Global Schema)				
Country	Site	Facility	Load	Timestamp
UK	RAL	CDF	0.3	19055711022002
UK	RAL	ATLAS	1.6	19055611022002
UK	GLA	CDF	0.4	19055811022002
UK	GLA	ALICE	0.5	19055611022002
CH	CERN	ALICE	0.9	19055611022002
CH	CERN	CDF	0.6	19055511022002

CPULoad (Producer 1)				
UK	RAL	CDF	0.3	19055711022002
UK	RAL	ATLAS	1.6	19055611022002

CPULoad (Producer 2)				
UK	GLA	CDF	0.4	19055811022002
UK	GLA	ALICE	0.5	19055611022002

CPULoad (Producer 3)				
CH	CERN	ATLAS	1.6	19055611022002
CH	CERN	CDF	0.6	19055511022002

Contributions are Views

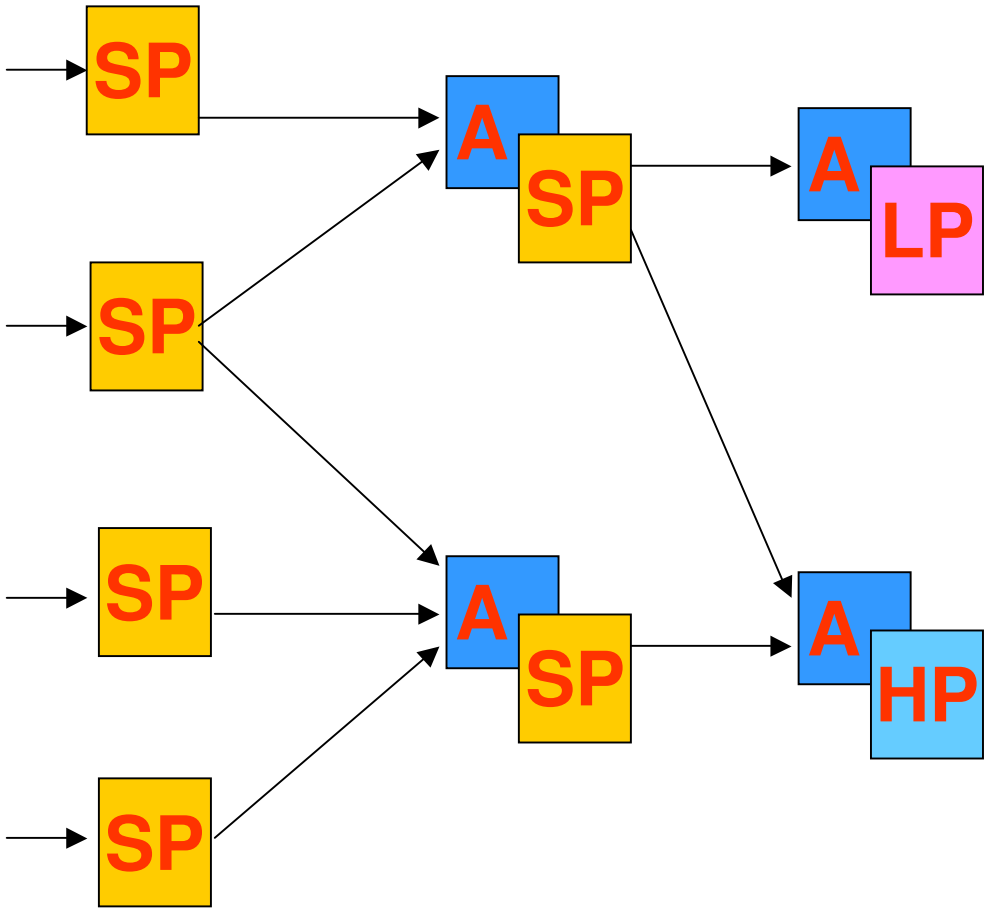
CPULoad (Producer 1)				
UK	RAL	CDF	0.3	19055711022002
UK	RAL	ATLAS	1.6	19055611022002

```
SELECT * FROM  cpuLoad

WHERE country = 'UK' AND site = 'RAL'
```

CPULoad (Producer 2)				
UK	GLA	CDF	0.4	19055811022002
UK	GLA	ALICE	0.5	19055611022002

```
SELECT * FROM  cpuLoad

WHERE country = 'UK' AND site = 'GLA'
```

Topologies



- Normally publish via a StreamProducer SP
- Archivers A instantiated with a Producer and a Predicate
- Must avoid cycles in the graph

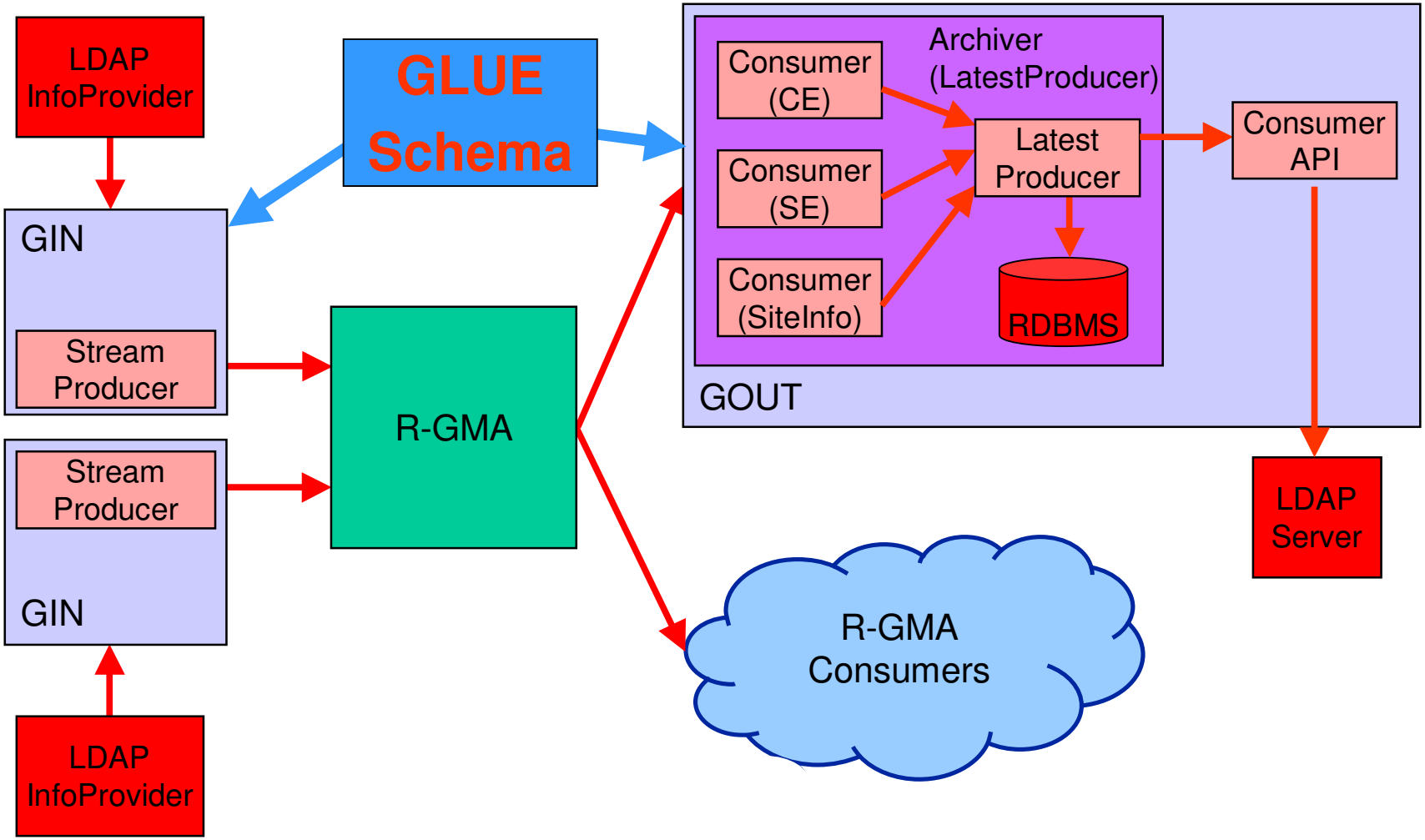
R-GMA Tools

- 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
- R-GMA Browser
  - JSP application dynamically generating web pages
  - Supports pre-defined and user-defined queries
- Pulse
  - R-GMA Java client-based GUI
  - Supports streaming and simple graphical displays

edg-rgma

- show tables
- describe Service
- show producers of Service
- latest select * from Service
- old continuous select * from Service

GIN and GOUT
(Gadget IN and Gadget OUT)



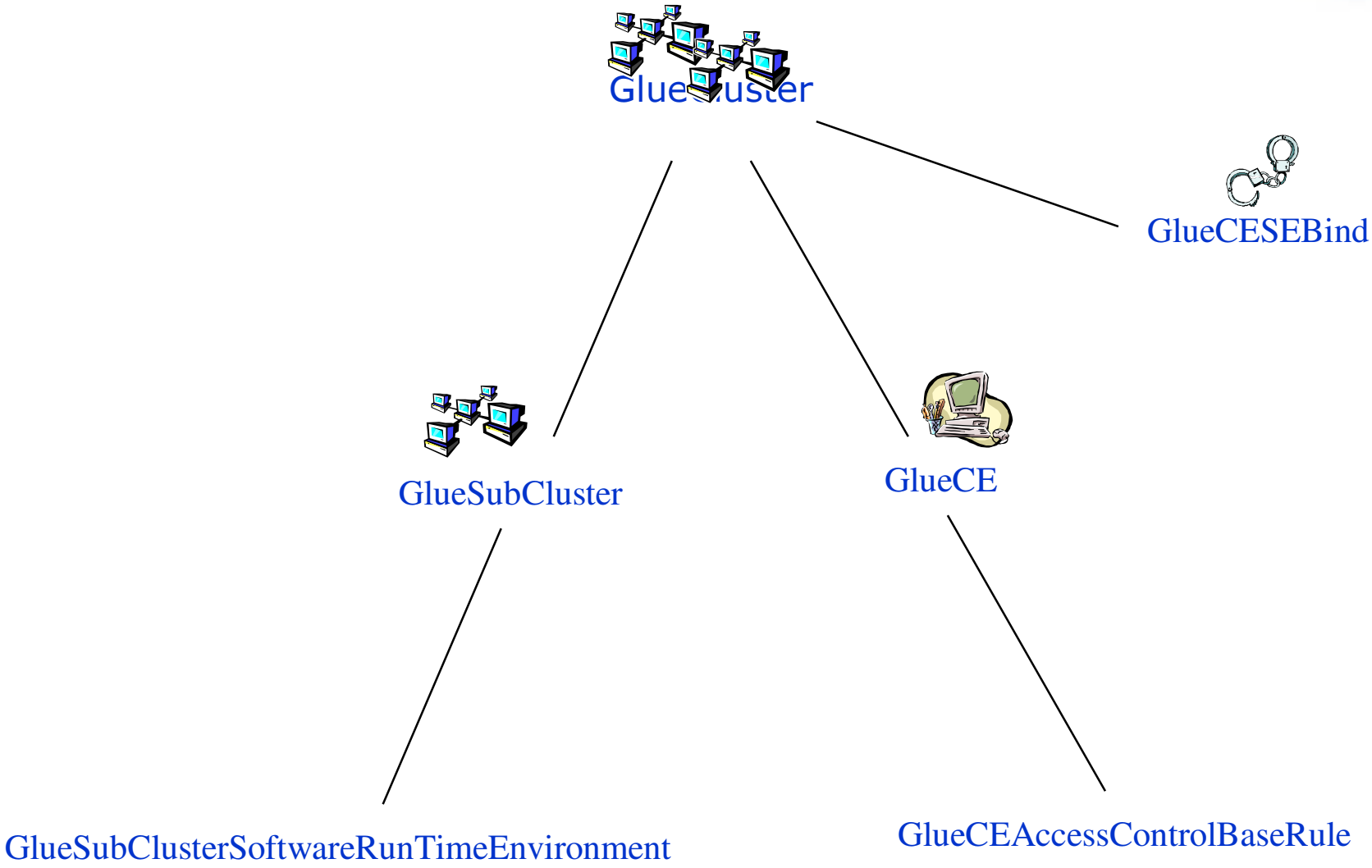
EDG Information Providers

- The information providers produce data in LDIF format
  - This is a legacy from when Globus MDS was the primary information system
- R-GMA publishes the data
  - Gin (gadget in) is used to invoke the information provider scripts
  - Gout (gadget out) republishes the data to a OpenLDAP database
    - This is to provide backwards compatibility during the transition from MDS to R-GMA
- EDG information providers
  - Software that provides information about resources and infrastructure
  - Provided by the work packages responsible for the resource

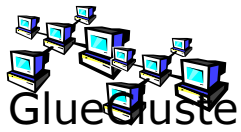
The GLUE schema

- Developed within HEP community
  - Globus
  - DataGrid
  - DataTAG
- It is a compromise - so not perfect for DataGrid
- Currently defines CEs and Ses
- Coming soon…
  - Network monitoring
  - Services

R-GMA Computing Service Tables



R-GMA Computing Service Tables



GlueCESEBind

UniqueID
Name
InformationServiceURL

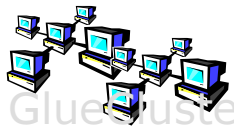
GlueSubCluster

GlueCE

GlueSubClusterSoftwareRunTimeEnvironment

GlueCEAccessControlBaseRule

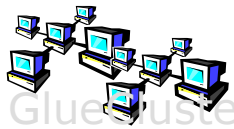
R-GMA Computing Service Tables



GlueSubCluster

UniqueID
Name
GlueClusterUniqueID
RAMSize
RAMAvailable
VirtualSize
VirtualAvailable
PlatformType
SMPSize
OSName
OSRelease
OSVersion
Vendor
Model
Version
ClockSpeed
InstructionSet
OtherProcessorDescription
CacheL1
CacheL1I
CacheL1D
CacheL2
BenchmarkSF00
BenchmarkSI00
InboundIP
OutboundIP
InformationServiceURL

R-GMA Computing Service Tables



GlueCluster

GlueCESEBind

GlueSubCluster

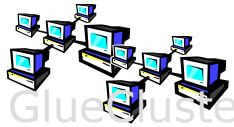
GlueCE

Value
GlueSubClusterUniqueID

GlueSubClusterSoftwareRunTimeEnvironment

GlueCEAccessControlBaseRule

R-GMA Computing Service Tables



GlueCluster

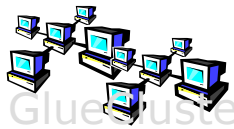
UniqueID
Name
GlueClusterUniqueID
TotalCPUs
LRMSType
LRMSVersion
GRAMVersion
HostName
GatekeeperPort
RunningJobs
WaitingJobs
TotalJobs
Status
WorstResponseTime
EstimatedResponseTime
FreeCpus
Priority
MaxRunningJobs
MaxTotalJobs
MaxCPUTime
MaxWallClockTime
InformationServiceURL

GlueCESEBind

GlueCE

GlueCEAccessControlBaseRule

R-GMA Computing Service Tables



GlueCluster

GlueCESEBind

GlueSubCluster

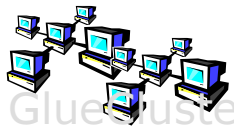
GlueCE

GlueCEUniqueID
Value

GlueCEAccessControlBaseRule

GlueSubClusterSoftwareRunTimeEnvironment

R-GMA Computing Service Tables



GlueCluster

GlueCESEBind

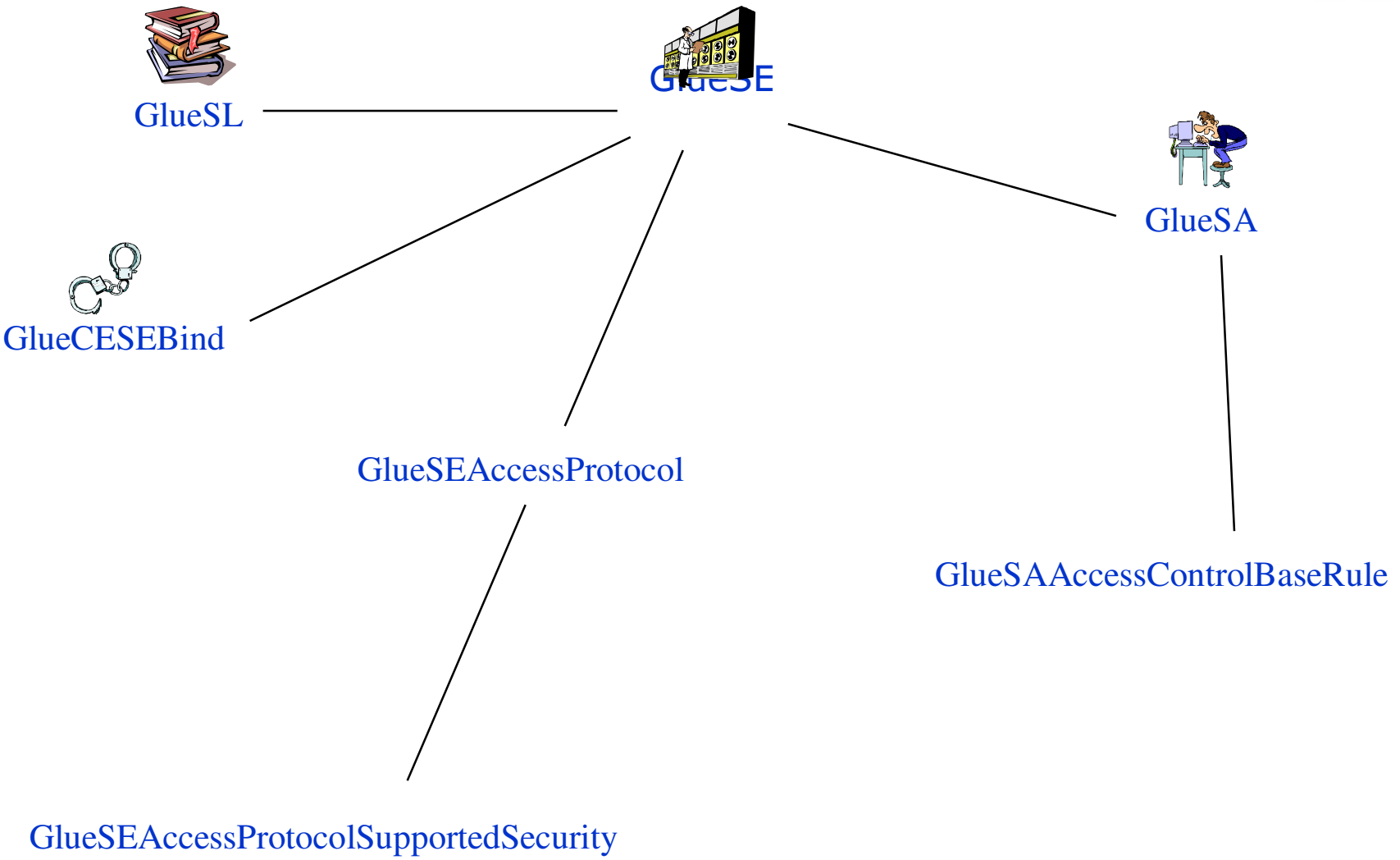
GlueSubCluster

GlueCEUniqueID
GlueSEUniqueID
Accesspoint

GlueSubClusterSoftwareRunTimeEnvironment

GlueCEAccessControlBaseRule

R-GMA Storage Service Tables



R-GMA Storage Service Tables

GlueSL

GlueSE

GlueSA

GlueCESEBind

GlueCEUniqueID
GlueSEUniqueID
Accesspoint

GlueSEAccessProtocol

GlueSAAccessControlBaseRule

GlueSEAccessProtocolSupportedSecurity

R-GMA Storage Service Tables

GlueSL

GlueSE

GlueSA

GlueCESEBind

GlueSEAccessProtocol

GlueSEAccessProtocolSupportedSecurity

UniqueID
GlueSLUniqueID
Name
Port
CurrentIOLoad
InformationServiceURL

R-GMA Storage Service Tables

GlueSL

GlueSE

GlueSA

GlueCESEBind

UniqueID
GlueSEUniqueID
Name
ArchitectureType
MaxIOCapacity
InformationServiceURL

GlueSAAccessControlBaseRule

GlueSEAccessProtocolSupportedSecurity

R-GMA Storage Service Tables

GlueSL

GlueSE

GlueSA

GlueCESEBind

GlueSEAccessProtocol

GlueSAAccessControlBaseRule

GlueSEAccessProtocolSupportedSecurity

UniqueID
GlueSEUniqueID
Type
Port
Version
AccessTime

R-GMA Storage Service Tables

GlueSL

GlueSE

GlueSA

GlueCESEBind

GlueSEAccessProtocol

GlueSEAccessProtocolUniqueID
value

GlueSEAccessProtocolSupportedSecurity

R-GMA Storage Service Tables

GlueSL

GlueSE

GlueSA

GlueCESEBind

GlueSEAcces

eSAAccessControlBaseRule

GlueSEAccessProtocolSupportedSecurity

UniqueID
Root
GlueSEUniqueID
MaxFileSize
MinFileSize
MaxData
MaxNumFiles
MaxPinDuration
Quota
FileLifeTime
AvailableSpace
UsedSpace

R-GMA Storage Service Tables

GlueSL

GlueSE

GlueSA

GlueCESEBind

GlueSEAccessProtocol

GlueSAAccessControlBaseRule

GlueSAUniqueID
Value

GlueSEAccessProtocolSupportedSecurity

Further Information

- Information and Monitoring Services
  - http://hepunx.rl.ac.uk/edg/wp3/