eGee

Enabling Grids for E-sciencE

gLite Information System: R-GMA

Tony Calanducci

INFN Catania

gLite tutorial at the EGEE User Forum

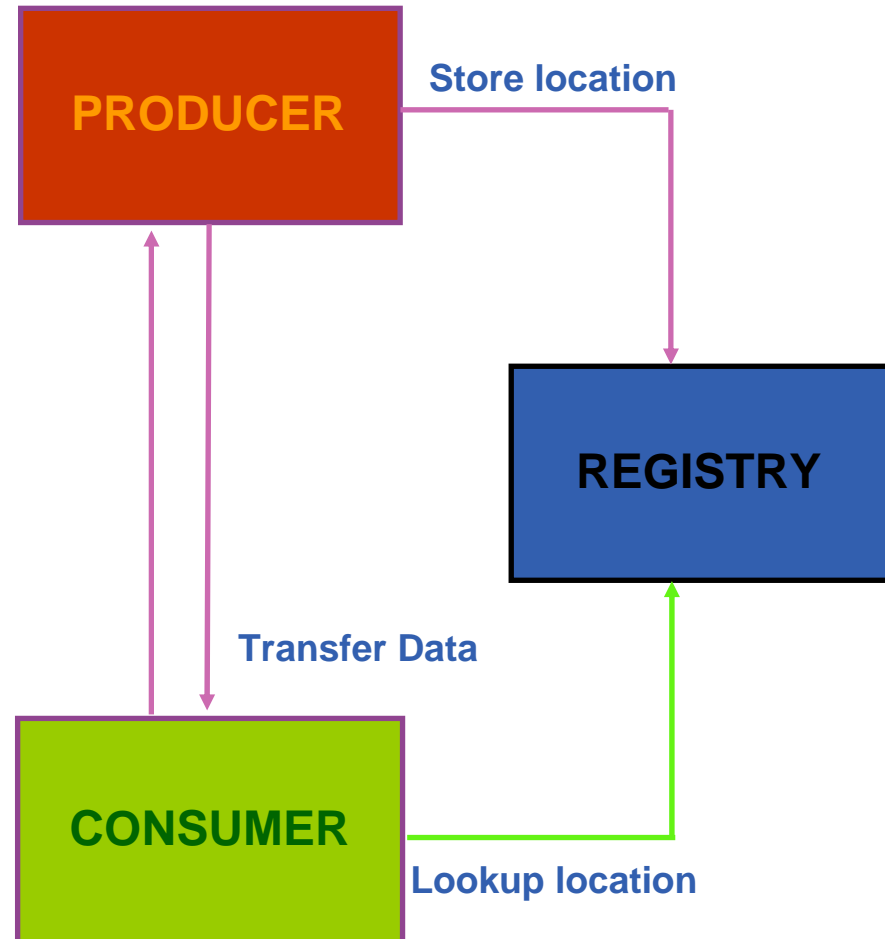*Geneva, 27-28 February 2006*

www.eu-egee.org

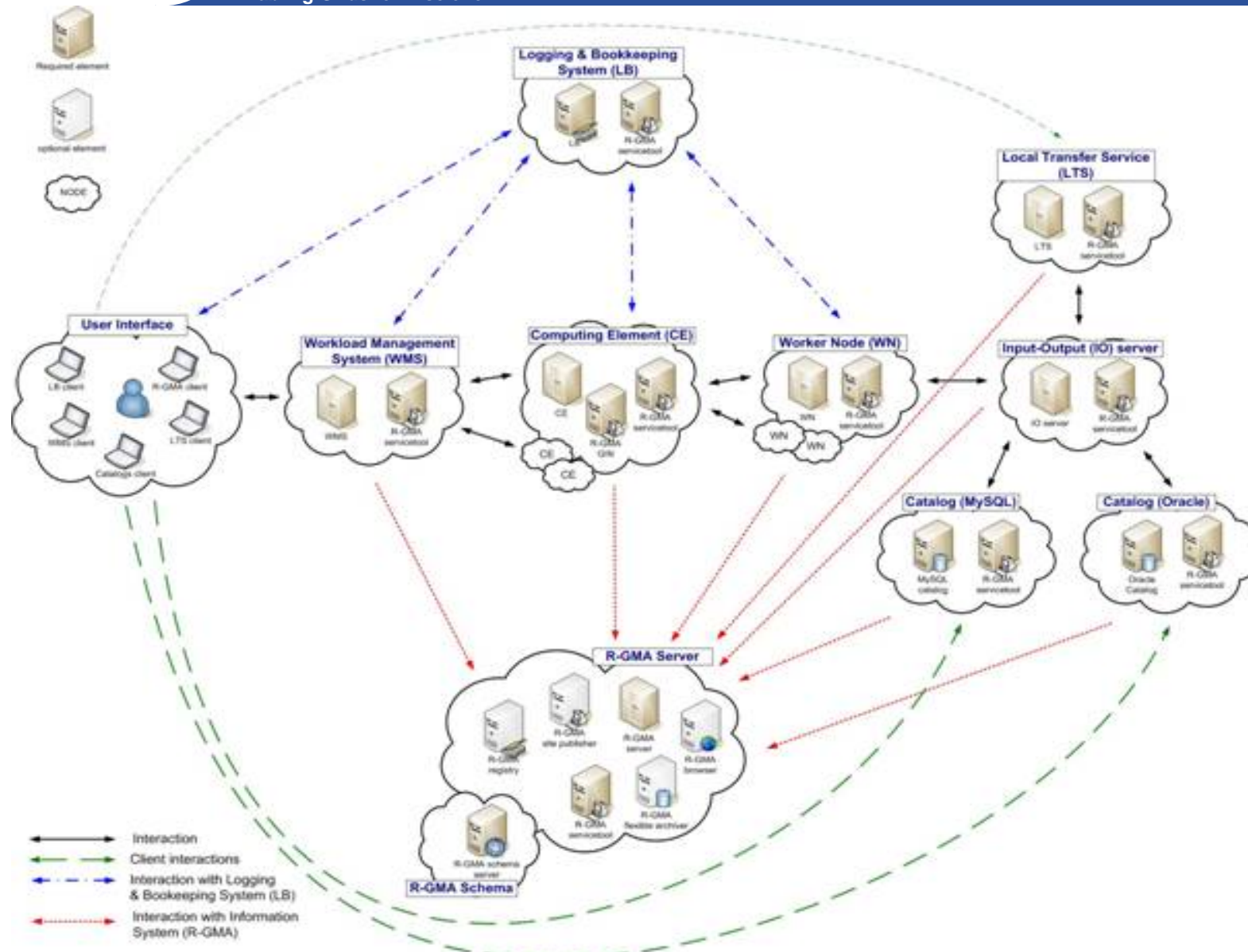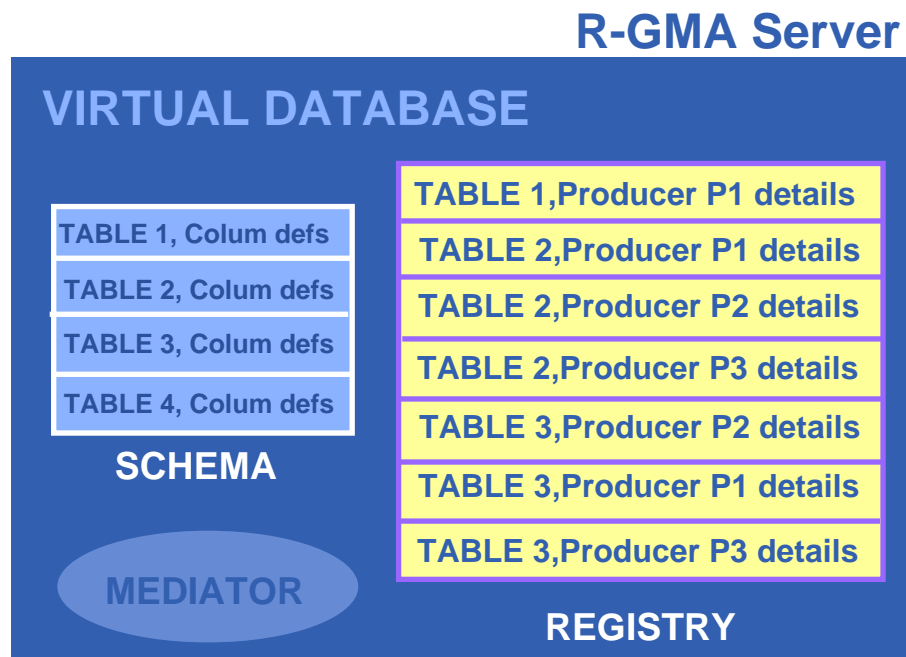Information Society
and Media

Enabling Grids for E-sciencE

➢ **Introduction to R-GMA and Grid Monitoring Architecture (GMA).**

➢ **R-GMA within testbeds**

➢ **R-GMA in depth:**
 - **Schema, Registry, Producer(s) and Consumer(s)**
 - **Query and Storage Types**
 - **R-GMA Browser**

➢ **References**

Enabling Grids for E-sciencE

- **Relational Grid Monitoring Architecture (R-GMA)**
  - Developed as part of the European DataGrid Project (EDG)
  - Now as part of the EGEE project.
  - Based on the Grid Monitoring Architecture (GMA) endorsed by the Global Grid Forum (GGF).

- **Uses a relational-like data model.**
  - Data are viewed as db tables.
  - Data structure defined by table columns.
  - Each entry is a row (tuple).
  - Queried using Structured Query Language (SQL).

- **The Producer stores its location (URL) in the Registry.**

- **The Consumer looks up producer URLs in the Registry.**

- **The Consumer contacts the Producer to get all the data or the Consumer can listen to the Producer for new data.**

PRODUCER

REGISTRY

CONSUMER

Store location

Transfer Data

Lookup location

Enabling Grids for E-sciencE

R-GMA Server

VIRTUAL DATABASE

TABLE 1, Colum defs

TABLE 2, Colum defs

TABLE 3, Colum defs

TABLE 4, Colum defs

SCHEMA

MEDIATOR

TABLE 1,Producer P1 details

TABLE 2,Producer P1 details

TABLE 2,Producer P2 details

TABLE 2,Producer P3 details

TABLE 3,Producer P2 details

TABLE 3,Producer P1 details
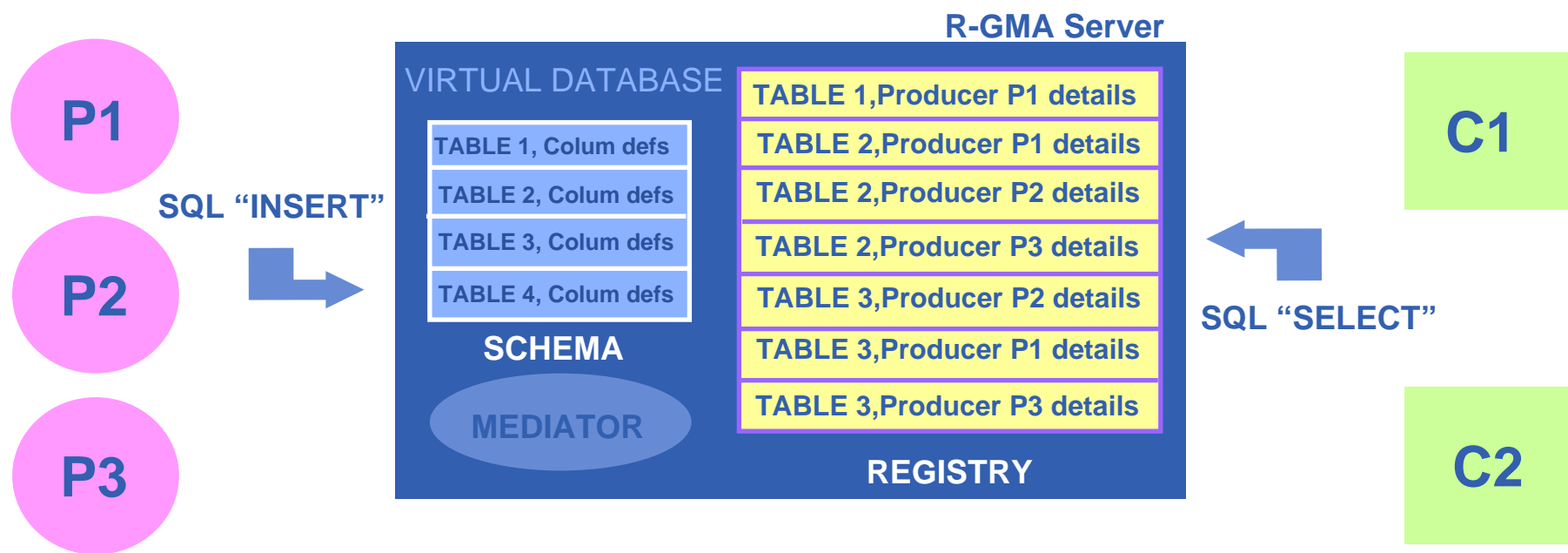
TABLE 3,Producer P3 details

REGISTRY

SCHEMA : it holds the names and definitions of all of the tables in the virtual database, and their authorization rules.

REGISTRY: It holds the details of all producers that are publishing to tables in the virtual database and it also holds the details of "continuous" consumers.
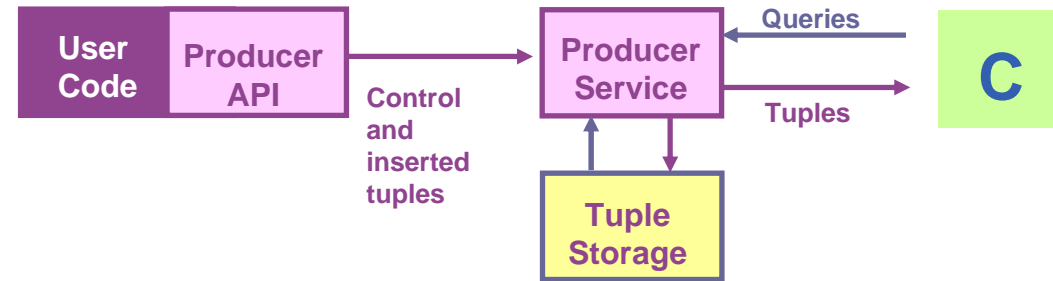
MEDIATOR: a set of rules for deciding which data providers to contact for any given query.

Producers: are the data providers for the virtual database. Writing data into the virtual database is known as publishing, and data is always published in complete rows, known as tuples. There are three types of producer: Primary, Secondary and On-demand.
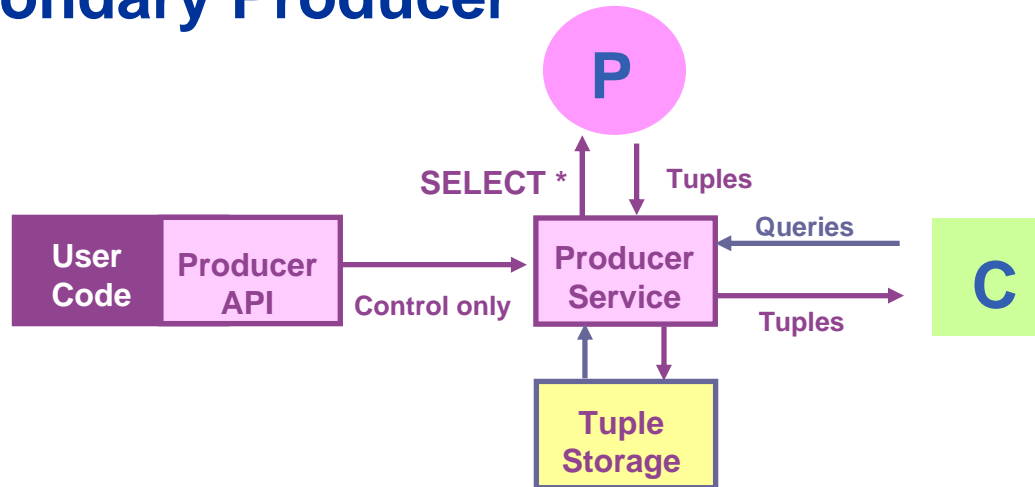
R-GMA Server

P1

P2

P3

SQL "INSERT"

VIRTUAL DATABASE

| TABLE 1, Colum defs |
| TABLE 2, Colum defs |
| TABLE 3, Colum defs |
| TABLE 4, Colum defs |

SCHEMA

MEDIATOR

| TABLE 1,Producer P1 details |
| TABLE 2,Producer P1 details |
| TABLE 2,Producer P2 details |
| TABLE 2,Producer P3 details |
| TABLE 3,Producer P2 details |
| TABLE 3,Producer P1 details |
| TABLE 3,Producer P3 details |

REGISTRY

C1

SQL "SELECT"

C2

Consumer: represents a single SQL SELECT query on the virtual database. The query is matched against the list of available producers in the Registry. The consumer service then selects the best set of producers to contact and sends the query directly to each of them, to obtain the answer tuples.
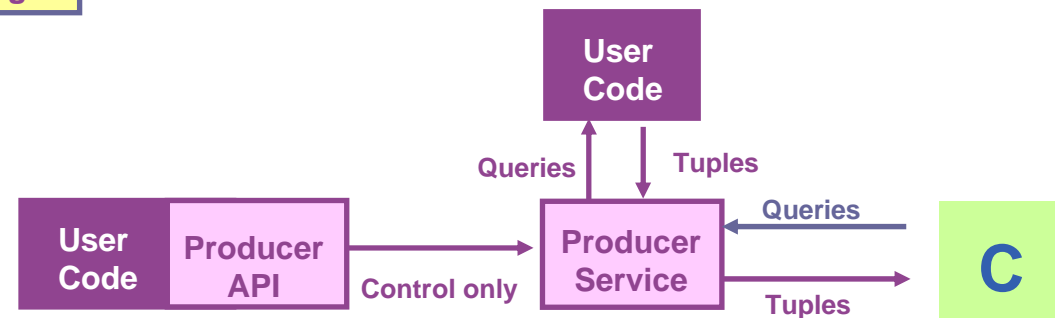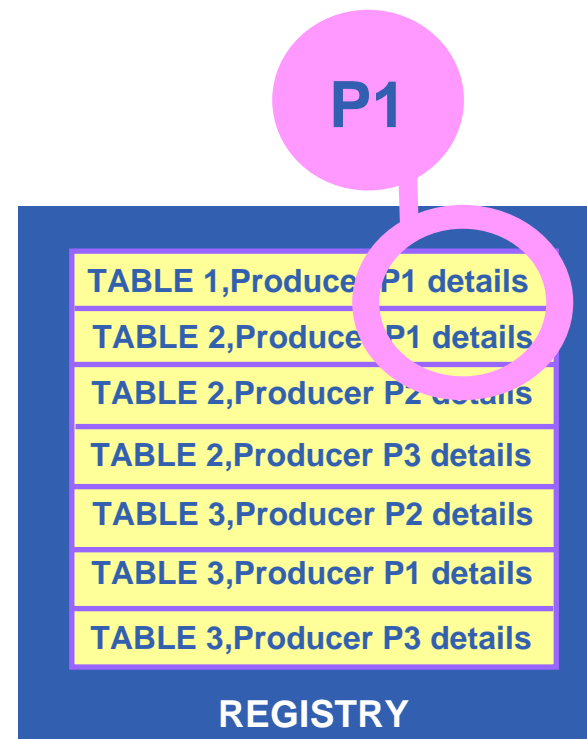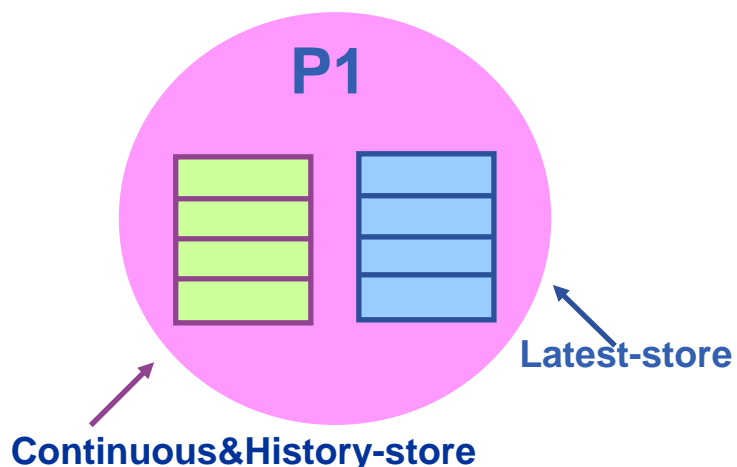
Enabling Grids for E-sciencE

- ## Primary Producer

| User Code | Producer API | | Producer Service | | C |

Control and inserted tuples

Queries

Tuples

Tuple Storage

- ## Secondary Producer

P

SELECT * *Tuples*

| User Code | Producer API | | Producer Service | | C |

Control only

Queries

Tuples

Tuple Storage

- ## On-Demand Producer

User Code

Queries *Tuples*

| User Code | Producer API | | Producer Service | | C |

Control only

Queries

Tuples

Enabling Grids for E-sciencE

- **Continuous:** as soon as new data becomes available it is broadcast to all interested parties.
- **Latest:** correspond to intuitive idea of current information.
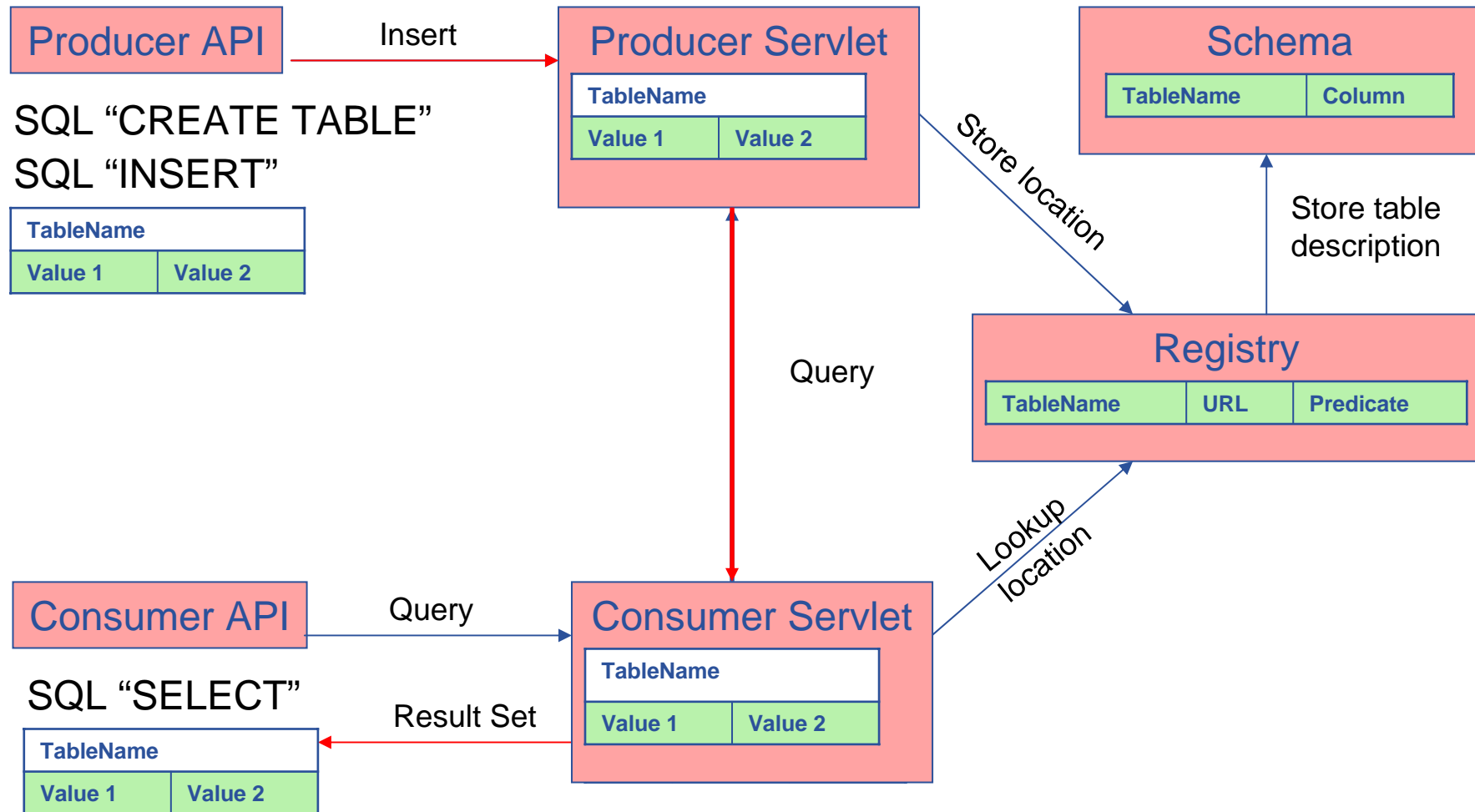- **History:** return time sequenced data.

Tuple-store can be in Memory or Database

P1

TABLE 1,Producer P1 details
TABLE 2,Producer P1 details
TABLE 2,Producer P2 details
TABLE 2,Producer P3 details
TABLE 3,Producer P2 details
TABLE 3,Producer P1 details
TABLE 3,Producer P3 details

REGISTRY

P1

Latest-store

Continuous&History-store

LATEST RETENTION PERIOD (LRP) and

HISTORY RETENTION PERIOD (RTP)

allow producers to periodically purge old tuples, and to give a precise meaning to the "current state".

Producer API —Insert→ **Producer Servlet**

SQL "CREATE TABLE"
SQL "INSERT"

TableName	
Value 1	Value 2

Producer Servlet

TableName	
Value 1	Value 2

Schema

TableName	Column

Store location

Store table description

Continuous

Registry

TableName	URL	Predicate

Lookup location

Consumer API —Query→ **Consumer Servlet**

SQL "SELECT"

TableName	
Value 1	Value 2

Result Set

Consumer Servlet

TableName	
Value 1	Value 2

Enabling Grids for E-sciencE

Producer API

SQL "CREATE TABLE"
SQL "INSERT"

TableName	
Value 1	Value 2

Insert →

Producer Servlet	
TableName	
Value 1	Value 2

Store location →

Schema	
TableName	**Column**

Store table description

Query

Registry		
TableName	**URL**	**Predicate**

Lookup location

Consumer API

SQL "SELECT"

TableName	
Value 1	Value 2

Query →

Consumer Servlet	
TableName	
Value 1	Value 2

Result Set

Enabling Grids for E-sciencE

• **Consumer users: want to be able to run queries on IS of any VO they belong.**

• **Producer users: want to be able to publish data to the IS of one or more VOs.**

• **Site administrators: want to control who creates R-GMA services on their systems and which services are run.**

• **Virtual Organizations: want to control who can add virtual tables to their schema and who can remove them.**

Producer   Consumer

R-GMA

VO   Site Admin

Consumer

R-GMA

Site Admin

Provider

VO
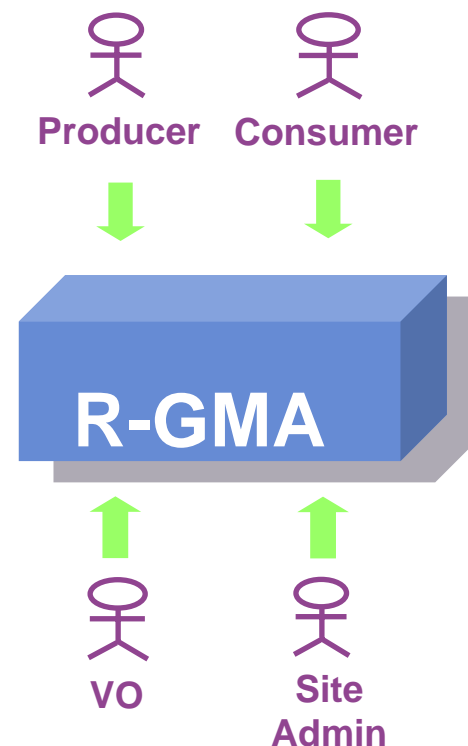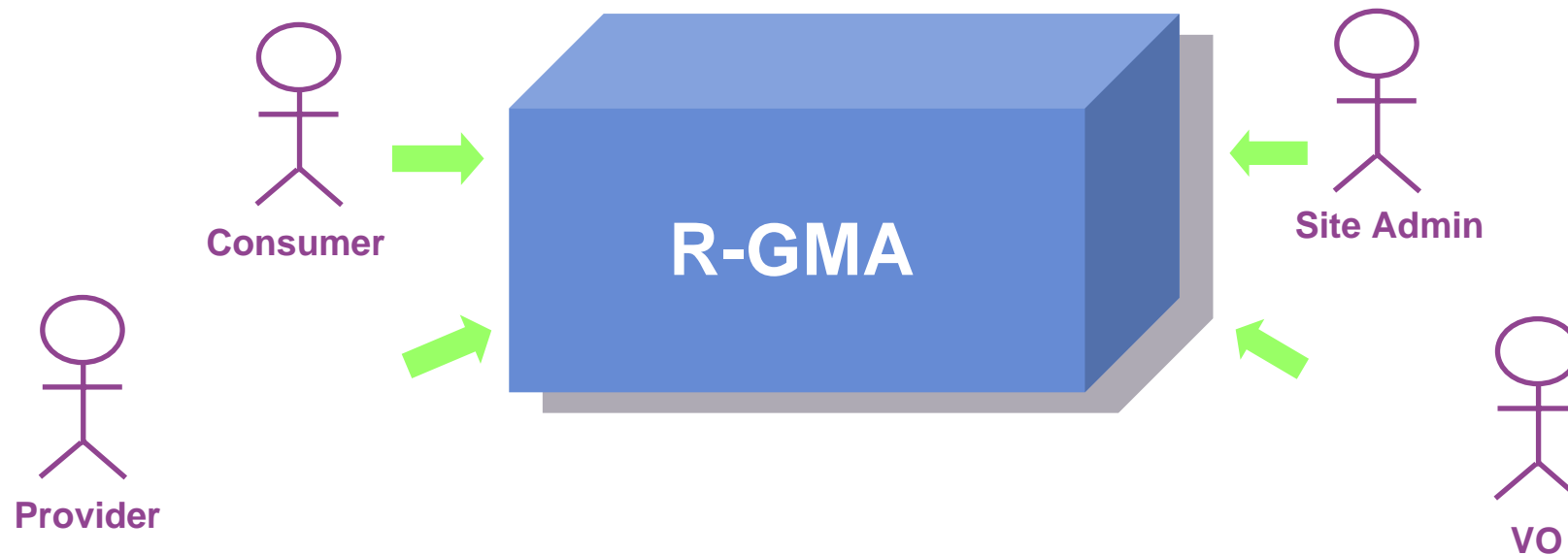
- **Mutual Autentication: guaranteeing who is at each end of an exchange of messages.**

- **Encryption: using an encrypted transport protocol (HTTPS).**

- **Authorization: it must be mutual. It can be implicit or explict.**

Enabling Grids for E-sciencE

- **R-GMA overview page.**
  - http://www.r-gma.org/

- **R-GMA documentation in EGEE**
  - http://hepunx.rl.ac.uk/egee/jra1-uk/

- **R-GMA browser in GILDA**
  - https://rgmasrv.ct.infn.it:8443/R-GMA