Information System

Giuseppe La Rocca
Valeria Ardizzone
INFN

EGEE Tutorial
Rome, 02-04 November 2005

lcg-infosites
&
lcg-info

How to discover resources ?

- Once an user is logged into an User Interface (s)he is ready to take advantage of the Grid Power for his/her own application.
- But what are the available resources to accomplish his/her tasks?
- The answer to this question comes through the interactions with the Information System (IS).
- The Information System (IS) provides information about the LCG-2 Grid resources and their status.

How to discover resources (cont)

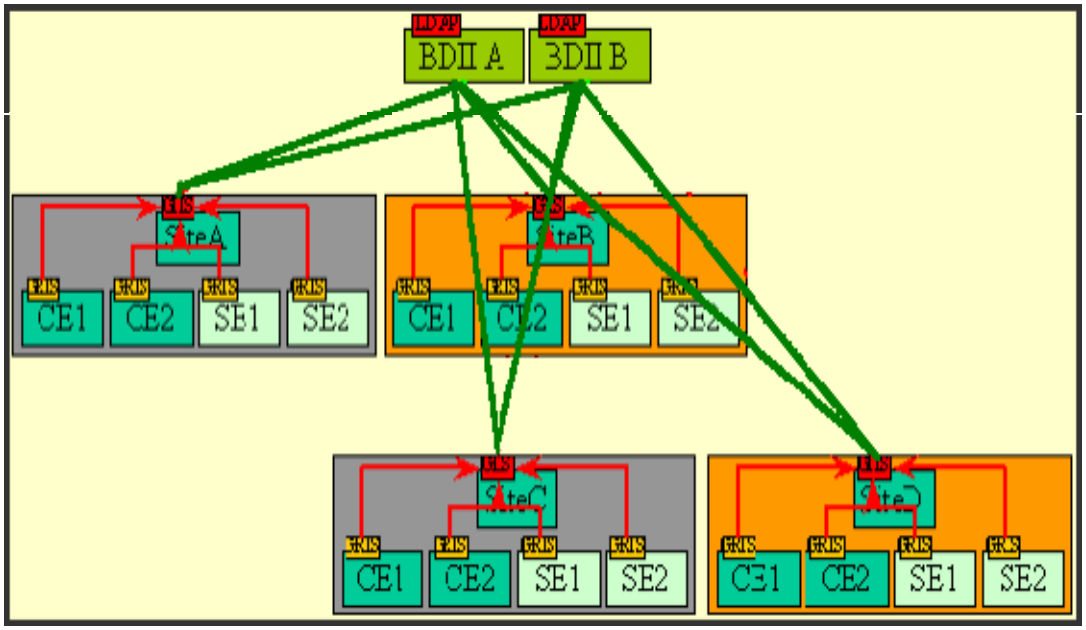
- The data published in the IS conforms to the GLUE (Grid Laboratory for a Uniform Environment) Schema. The GLUE Schema aims to define a common conceptual data model to be used for Grid resources.

- In LCG-2, the BDII (Berkeley DB Information Index), based on an updated version of the Monitoring and Discovery Service (MDS), was adopted as main provider of the Information Service.

- In gLite, R-GMA (Relational Grid Monitoring Architecture) is adopted as IS.

Monitoring and Discovery Service

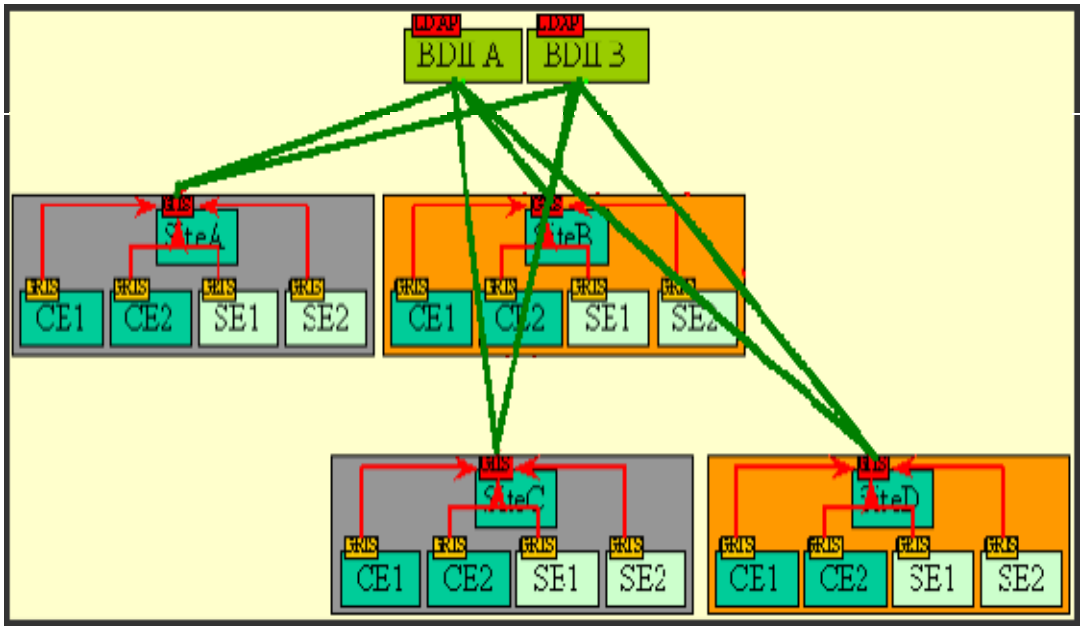
- Computing and storage resources at a site implement an entity called Information Provider, which generates the relevant information of the resource (e.g.: the used space in a SE).
- This information is published via an LDAP server by the Grid Resource Information Servers, or GRISes.



Monitoring and Discovery Service

- In each site an element called the Site Grid Index Information Server (GIIS) collects all the information of the different GRISes and publishes it.

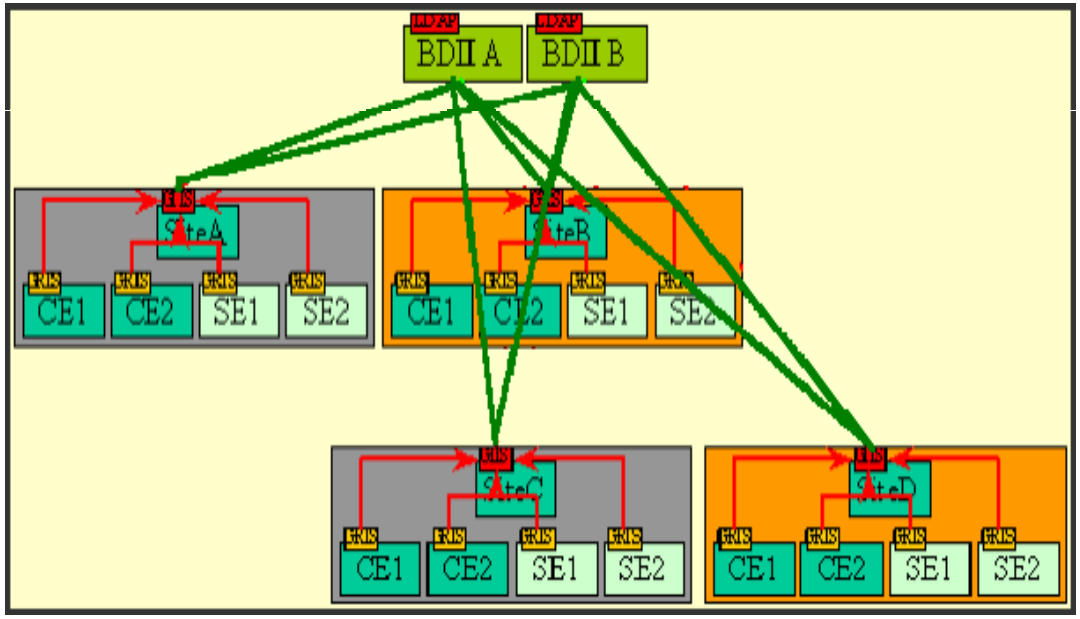
- This BDII queries the GIISes and acts as a cache, storing information about the Grid status in its database.



Monitoring and Discovery Service

- Querying the BDII a user or a service has all the available information about the status of the grid resources.

- Moreover in order to get more up-to-date information it is possible to querying directly the GIISes or GRISes.



How to query the IS?

- In order to query directly the IS elements two higher level tools are provided.

lcg-infosites

lcg-info

- These tools should be enough for most common user needs and will usually avoid the necessary of raw LDAP queries.

lcg-infosites

- The lcg-infosites command can be used as an easy way to retrieve information on Grid resources for the most use cases.

USAGE: lcg-infosites --vo <vo name> options -v <verbose level> --is <BDII to query>

lcg-infosites options

ce	The information related to number of CPUs, running jobs, waiting jobs and names of the CEs are provided. All these data group all VOs together. With "-v 1" only the names of the queues will be printed while with "-v 2" The RAM Memory together with the operating system and its version and the processor included in each CE are printed.
se	The names of the SEs supported by the user's VO together with the kind of Storage System,the used and available space will be printed. With "-v 1" only the names of the SEs will be printed.
closeSE	The names of the CEs where the user's VO is allowed to run together with their corresponding closest SEs are provided.
lfc	Name of the lfc Catalog for the user's VO.
tag	The names of the tags relative to the software installed in site is printed together with the corresponding CE.
all	It groups together the information provided by ce, se, lrc and rmc.
is	If not specified the BDII defined in default by the variable LCG GFAL INFOSYS will be queries. However the user may want to query any other BDII without redefining this environment variable. This is possible specifying this argument followed by the name of the BDII which the user wants to query. All options admits this argument.

Obtaining information about CE

Enabling Grids for E-sciencE

$ lcg-infosites --vo gilda ce

```
*****************************************************************
These are the related data for gilda: (in terms of queues and CPUs)
*****************************************************************
#CPU   Free   Total Jobs    Running Waiting ComputingElement
-----------------------------------------------------------------------------------
  4      3       0            0       0        cn01.be.itu.edu.tr:2119/jobmanager-lcglsf-long
  4      3       0            0       0        cn01.be.itu.edu.tr:2119/jobmanager-lcglsf-short
 34     33       0            0       0        grid010.ct.infn.it:2119/jobmanager-lcgpbs-long
 16     16       0            0       0        grid011f.cnaf.infn.it:2119/jobmanager-lcgpbs-long
  1      1       0            0       0        grid006.cecalc.ula.ve:2119/jobmanager-lcgpbs-log
  2      1       1            0       1        gildace.oact.inaf.it:2119/jobmanager-lcgpbs-short
[..]
```

$ lcg-infosites --vo gilda ce --v 2

```
RAMMemory   Operating System   System Version   Processor   CE Name
-----------------------------------------------------------------------------------------------------
1024        SLC    3          P4        ced-ce0.datagrid.cnr.it
4096        SLC    3          Xeon      cn01.be.itu.edu.tr
1024        SLC    3          PIII      cna02.cna.unicamp.br
 917        SLC    3          PIII      gilda-ce-01.pd.infn.it
1024        SLC    3          Athlon    gildace.oact.inaf.it
1024        SLC    3          Xeon      grid-ce.bio.dist.unige.it
[..]
```

Obtaining information about SE

$ lcg-infosites --vo gilda se

****

These are the related data for gilda: (in terms of SE)

****

Avail Space(Kb)	Used Space(Kb)	Type	SEs
143547680	2472756	disk	cn02.be.itu.edu.tr
168727984	118549624	disk	grid009.ct.infn.it
13908644	2819288	disk	grid003.cecalc.ula.ve
108741124	2442872	disk	gildase.oact.inaf.it
28211488	2948292	disk	testbed005.cnaf.infn.it
349001680	33028	disk	gilda-se-01.pd.infn.it
31724384	2819596	disk	cna03.cna.unicamp.br
387834656	629136	disk	grid-se.bio.dist.unige.it

Listing the close Storage Elements

$ lcg-infosites --vo gilda closeSE

Name of the CE: cn01.be.itu.edu.tr:2119/jobmanager-lcglsf-long
Name of the close SE: cn02.be.itu.edu.tr

Name of the CE: cn01.be.itu.edu.tr:2119/jobmanager-lcglsf-short
Name of the close SE: cn02.be.itu.edu.tr

Name of the CE: grid010.ct.infn.it:2119/jobmanager-lcgpbs-long
Name of the close SE: grid009.ct.infn.it

Name of the CE: grid011f.cnaf.infn.it:2119/jobmanager-lcgpbs-long
Name of the close SE: testbed005.cnaf.infn.it

Listing tags of installed software

$ lcg-infosites --vo gilda tag

*****

lcg-info intro

- This command can be used to list either CEs or the SEs that satisfy a given set of conditions, and to print the values of a given set of attributes.
- The information is taken from the BDII specified by the LCG_GFAL_INFOSYS environment variable.

- The query syntax is like this:

```
attr1 op1 valueN, ... attrN opN valueN
```

where *attrN* is an attribute name

op is =, >= or <=, and the cuts are ANDed.

The cuts are comma-separated and spaces are not allowed.

lcg-info usage

USAGE

lcg-info --list-ce [--bdii bdii] [--vo vo] [--sed] [--query query] [--attrs list]

lcg-info --list-se [--bdii bdii] [--vo vo] [--sed] [--query query] [--attrs list]

lcg-info --list-attrs

lcg-info --help

lcg-info options

--list-attrs	Prints a list of the attributes that can be queried.
--list-ce	Lists the CEs which satisfy a query, or all the CEs if no query is given.
--list-se	Lists the SEs which satisfy a query, or all the SEs if no query is given.
--query	Restricts the output to the CEs (SEs) which satisfy the given query.
--bdii	Allows to specify a BDII in the form :. If not given, the value of the environmental variable LCG_GFAL_INFOSYS is used. If that is not defined, the command returns an error.
--sed	Print the output in a "sed-friendly" format.
--attrs	Specifies the attributes whose values should be printed.
--vo	Restricts the output to CEs or SEs where the given VO is authorized. Mandatory when VO-dependent attributes are queried upon.

Get the list of supported attributes

```
$ lcg-info --list-attrs
```

Attribute name	Glue object class	Glue attribute name
MaxTime	GlueCE	GlueCEPolicyMaxWallClockTime
CEStatus	GlueCE	GlueCEStateStatus
TotalJobs	GlueCE	GlueCEStateTotalJobs
CEVOs	GlueCE	GlueCEAccessControlBaseRule
TotalCPUs	GlueCE	GlueCEInfoTotalCPUs
FreeCPUs	GlueCE	GlueCEStateFreeCPUs
CE	GlueCE	GlueCEUniqueID
WaitingJobs	GlueCE	GlueCEStateWaitingJobs
RunningJobs	GlueCE	GlueCEStateRunningJobs
CloseCE	GlueCESEBindGroup	GlueCESEBindGroupCEUniqueID
CloseSE	GlueCESEBindGroup	GlueCESEBindGroupSEUniqueID
SEVOs	GlueSA	GlueSAAccessControlBaseRule
UsedSpace	GlueSA	GlueSAStateUsedSpace
AvailableSpace	GlueSA	GlueSAStateAvailableSpace
Type	GlueSE	GlueSEType
SE	GlueSE	GlueSEUniqueID
Protocol	GlueSEAccessProtocol	GlueSEAccessProtocolType
ArchType	GlueSL	GlueSLArchitectureType
Processor	GlueSubCluster	GlueHostProcessorModel
OS	GlueSubCluster	GlueHostOperatingSystemName
Cluster	GlueSubCluster	GlueSubClusterUniqueID
Tag	GlueSubCluster	GlueHostApplicationSoftwareRunTimeEnvironment
Memory	GlueSubCluster	GlueHostMainMemoryRAMSize

lcg-info examples

List all the CE(s) in the BDII satisfying given conditions

$ lcg-info --list-ce --query 'TotalCPUs>=30,OS=SL*' --attrs 'RunningJobs,FreeCPUs'

```
- CE: grid010.ct.infn.it:2119/jobmanager-lcgpbs-long
  - RunningJobs        0
  - FreeCPUs          33

- CE: grid010.ct.infn.it:2119/jobmanager-lcgpbs-short
  - RunningJobs        0
  - FreeCPUs          33

- CE: grid010.ct.infn.it:2119/jobmanager-lcgpbs-infinite
  - RunningJobs        1
  - FreeCPUs          33

- CE: skurut1.cesnet.cz:2119/jobmanager-lcgpbs-long
  - RunningJobs        0
  - FreeCPUs          26

- CE: skurut1.cesnet.cz:2119/jobmanager-lcgpbs-gilda
  - RunningJobs        0
  - FreeCPUs          26
[..]
```

lcg-info examples (cont.)

List all the CE(s) which satisfying the condition FreeCPU >=30

```
$ lcg-info --list-ce --query 'FreeCPUs >= 30‘--attrs ‘FreeCPUs’

- CE: grid010.ct.infn.it:2119/jobmanager-lcgpbs-long
  - FreeCPUs          33

- CE: grid010.ct.infn.it:2119/jobmanager-lcgpbs-short
  - FreeCPUs          33

- CE: grid010.ct.infn.it:2119/jobmanager-lcgpbs-infinite
  - FreeCPUs          33

[..]
```

Print all the tags published by a specific query

```
$ lcg-info --list-ce --query 'CE=*grid010.ct.infn.it:2119*' --attrs 'Tag'
```

PBS
INFN
CATANIA
LCG-2
LCG-2_1_0
LCG-2_1_1
LCG-2_2_0
LCG-2_3_0
LCG-2_3_1
LCG-2_4_0
R-GMA
AFS
CMS-1.1.0
ATLAS-6.0.4
GATE-1.0.0-3
LHCb-1.1.1
IDL-5.4
CMSIM-125
ALICE-4.01.00
ALIEN-1.32.14
POVRAY-3.5
DEMTOOLS-1.0
CMKIN-VALID
CMKIN-1.1.0
CMSIM-VALID
CSOUND-4.13
MPICH
VIRGO-1.0
CMS-OSCAR-2.4.5
LHCb_dbase_common-v3r1
GEANT4-6
VLC-0.7.2
EGEODE-1.0
RASTER3D
SCILAB-2.6
G95-3.5.0
MAGIC-6.19
CODESA3D-1.0
VO-gilda-slc3_ia32_gcc323
VO-gilda-CMKIN_5_1_1
VO-gilda-GEANT
VO-gilda-GKS05

List the CEs with a particular SW

$ lcg-info –vo gilda --list-ce --query 'Tag=*MPICH*' --attrs 'CE'

- CE: cn01.be.itu.edu.tr:2119/jobmanager-lcglsf-long
- CE cn01.be.itu.edu.tr:2119/jobmanager-lcglsf-long

- CE: cn01.be.itu.edu.tr:2119/jobmanager-lcglsf-short
- CE cn01.be.itu.edu.tr:2119/jobmanager-lcglsf-short

- CE: grid010.ct.infn.it:2119/jobmanager-lcgpbs-long
- CE grid010.ct.infn.it:2119/jobmanager-lcgpbs-long

- CE: grid011f.cnaf.infn.it:2119/jobmanager-lcgpbs-long
- CE grid011f.cnaf.infn.it:2119/jobmanager-lcgpbs-long

- CE: ced-ce0.datagrid.cnr.it:2119/jobmanager-lcgpbs-long
- CE ced-ce0.datagrid.cnr.it:2119/jobmanager-lcgpbs-long

[..]

List the SEs satisfying given query

$ lcg-info –vo gilda --list-se --query 'AvailableSpace>=100000' --attrs 'CloseCE'

- SE: cn02.be.itu.edu.tr
- CloseCE cn01.be.itu.edu.tr:2119/jobmanager-lcglsf-long
cn01.be.itu.edu.tr:2119/jobmanager-lcglsf-short
cn01.be.itu.edu.tr:2119/jobmanager-lcglsf-infinite

- SE: grid009.ct.infn.it
- CloseCE grid010.ct.infn.it:2119/jobmanager-lcgpbs-long
grid010.ct.infn.it:2119/jobmanager-lcgpbs-short
grid010.ct.infn.it:2119/jobmanager-lcgpbs-infinite

- SE: ced-se0.datagrid.cnr.it
- CloseCE ced-ce0.datagrid.cnr.it:2119/jobmanager-lcgpbs-long
ced-ce0.datagrid.cnr.it:2119/jobmanager-lcgpbs-short
ced-ce0.datagrid.cnr.it:2119/jobmanager-lcgpbs-infinite

- SE: grid003.cecablc.ula.ve
- CloseCE grid006.cecalc.ula.ve:2119/jobmanager-lcgpbs-cert
grid006.cecalc.ula.ve:2119/jobmanager-lcgpbs-long
grid006.cecalc.ula.ve:2119/jobmanager-lcgpbs-short
grid006.cecalc.ula.ve:2119/jobmanager-lcgpbs-infinite

[..]

R-GMA

Outline

This part covers the following arguments:

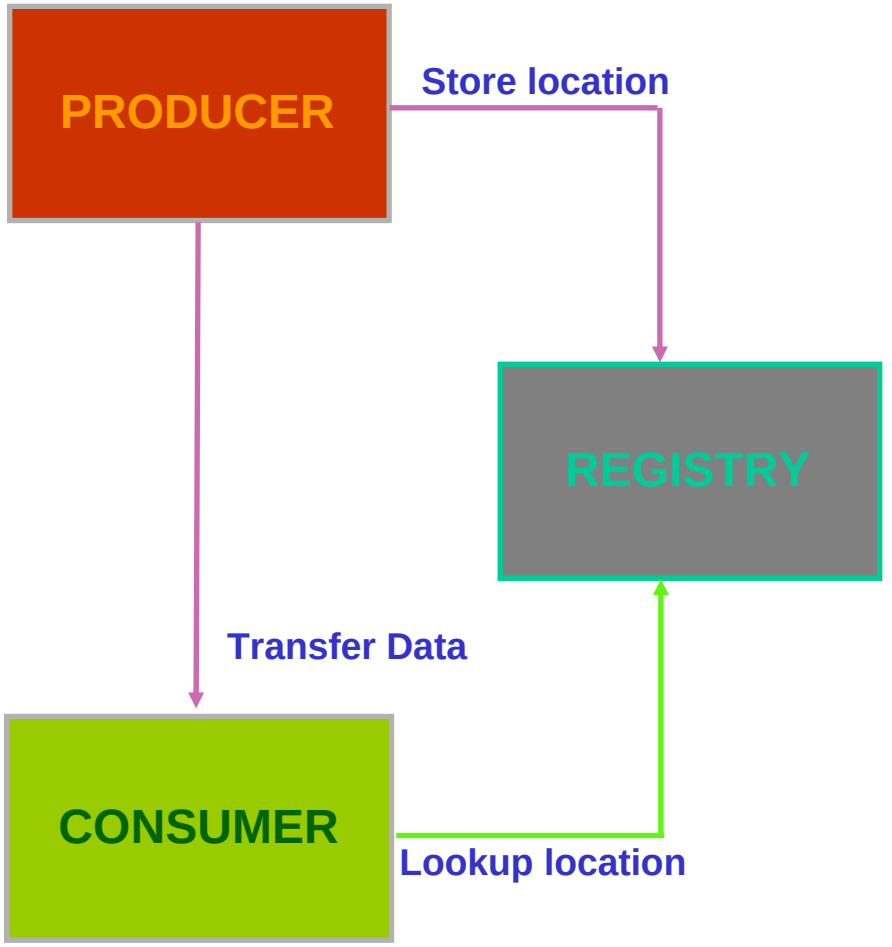
- Introduction to R-GMA.
- Grid Monitoring Architecture (GMA).
- R-GMA in depth:
  - Registry,Schema, Producer and Consumer
  - Query Type
  - R-GMA Browser
- Security in R-GMA.

Introduction to R-GMA

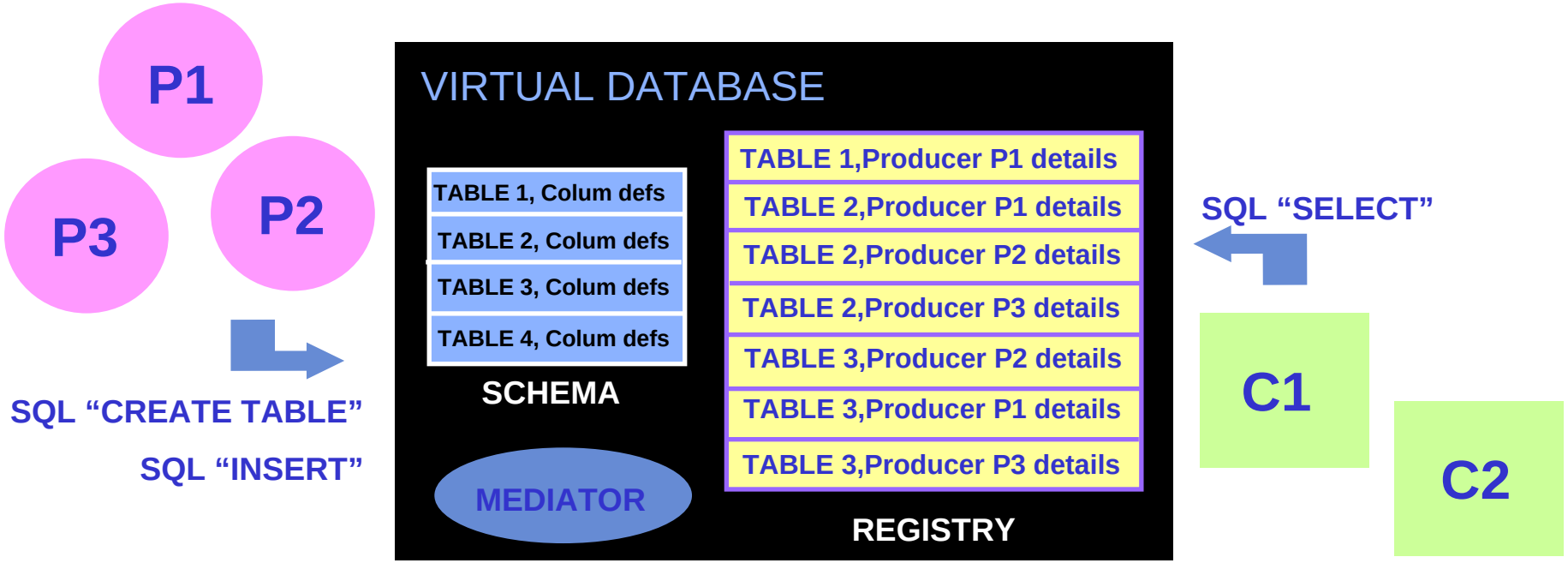
- **Relational Grid Monitoring Architecture (R-GMA)**
  - Developed as part of the EuropeanDataGrid Project (EDG)
  - Now as part of the EGEE project.
  - Based the Grid Monitoring Architecture (GMA) from the Global Grid Forum (GGF).
- **Uses a relational data model.**
  - Data are viewed as tables.
  - Data structure defined by the columns.
  - Each entry is a row (tuple).
  - Queried using Structured Query Language (SQL).

Grid Monitoring Architecture

- 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.



R-GMA in depth



There is no central repository!!! There is only a "*Virtual Database*".

Schema is a list of table definitions.

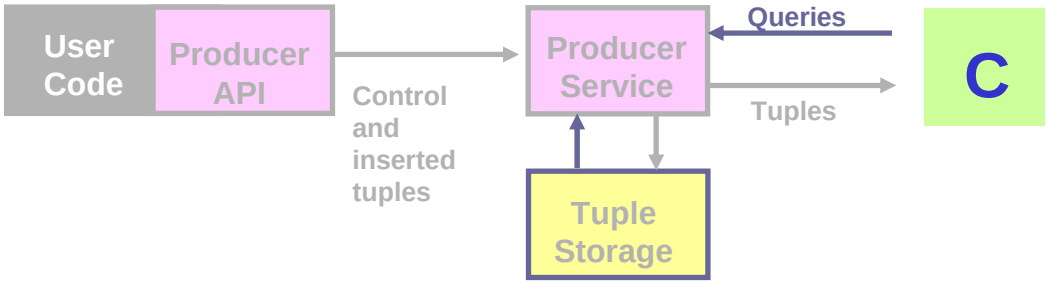
Registry is a list of data producers with all its details.

Producers publish data.

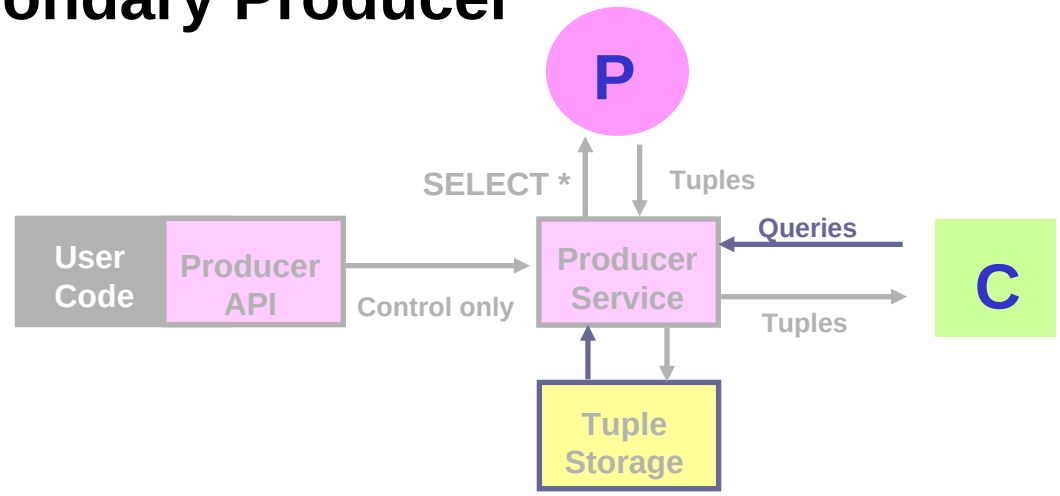
Consumer read data published.

Producer Types

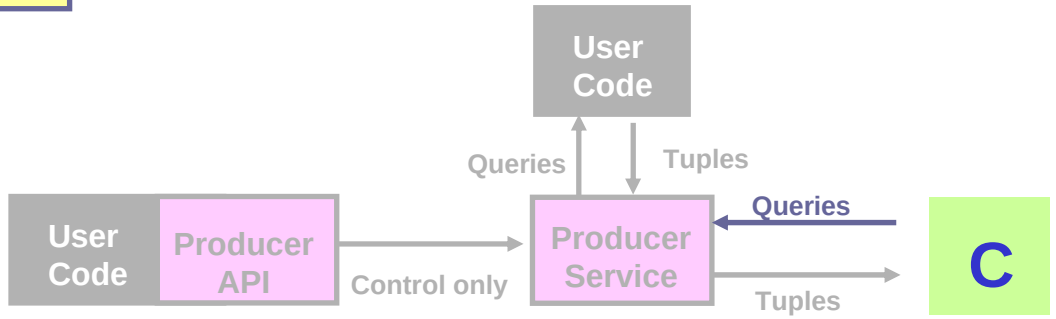
- **Primary Producer**



- **Secondary Producer**



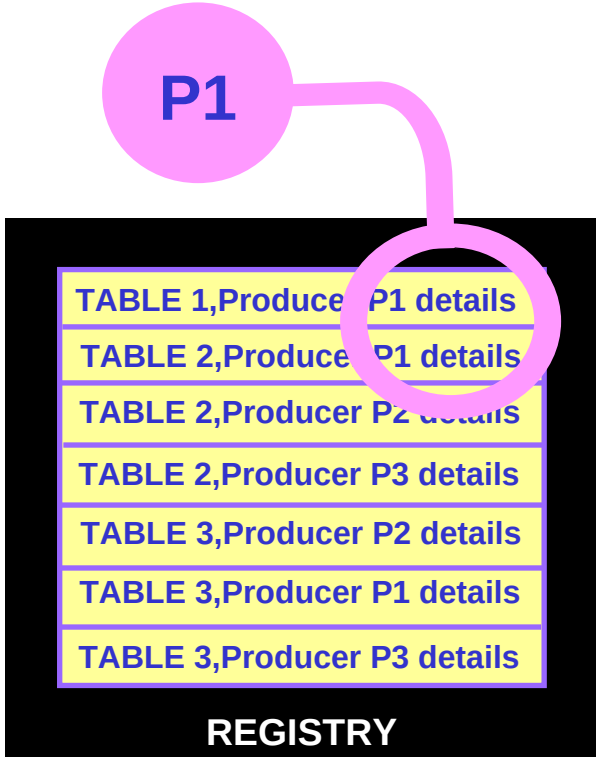
- **On-Demand Producer**



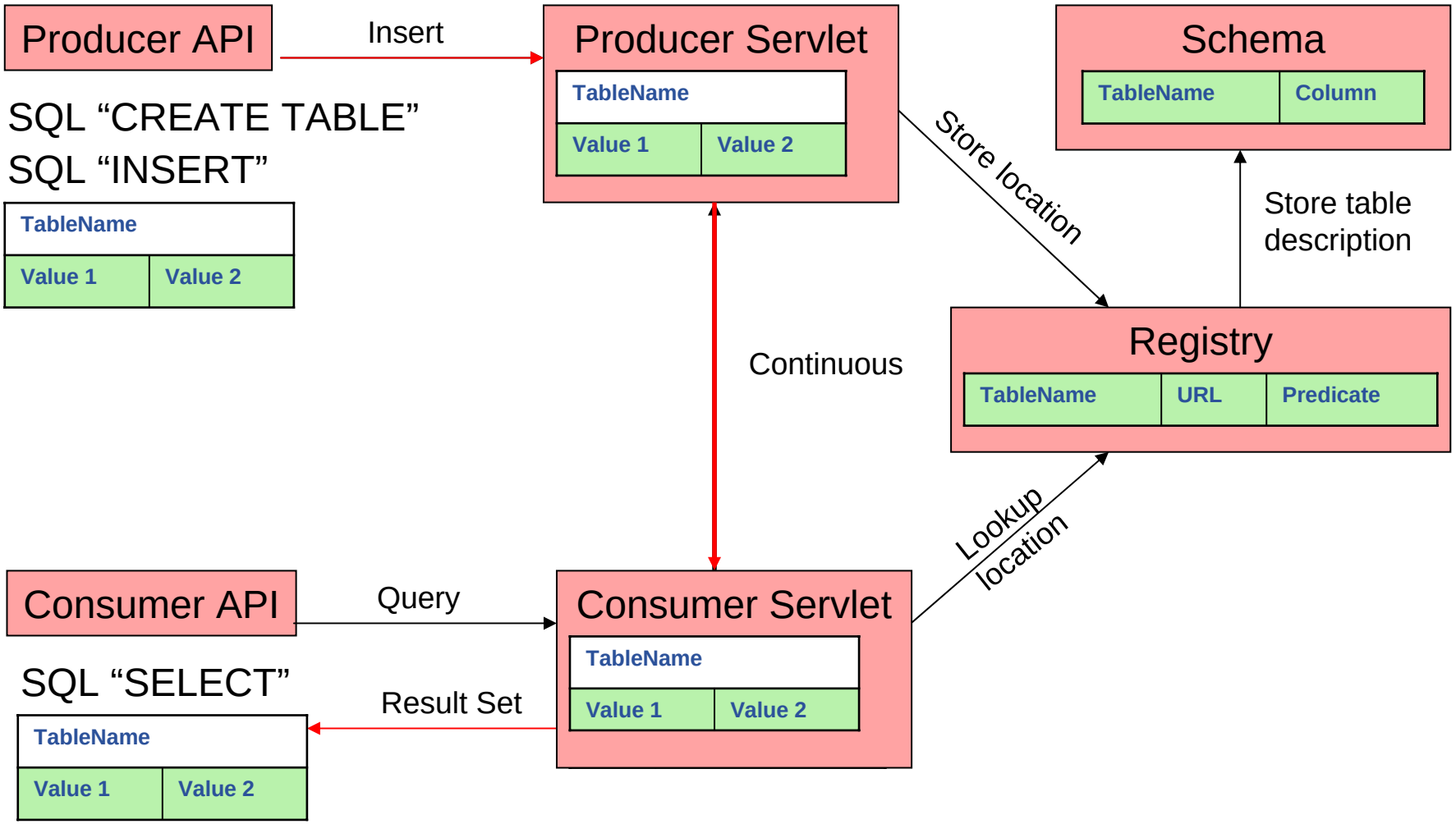
Query Types

Continuous

- **Latest**
- **History**
- **Static**

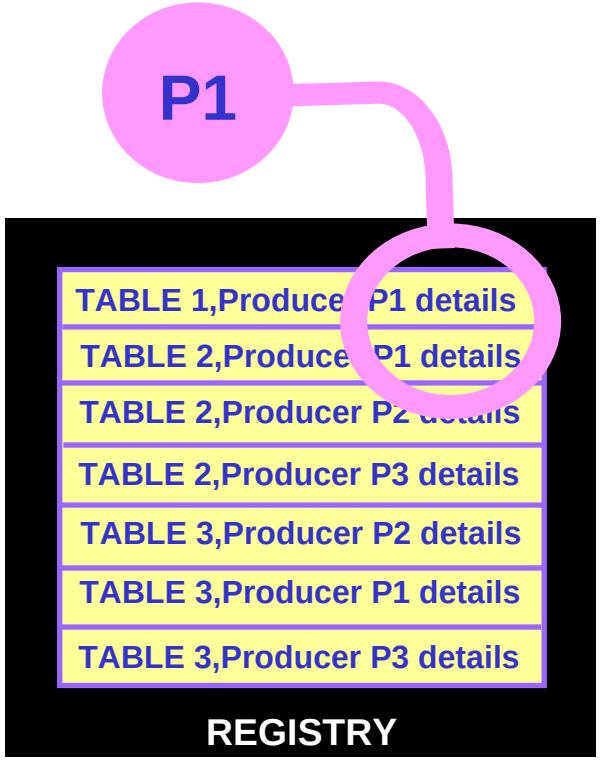


Continuous

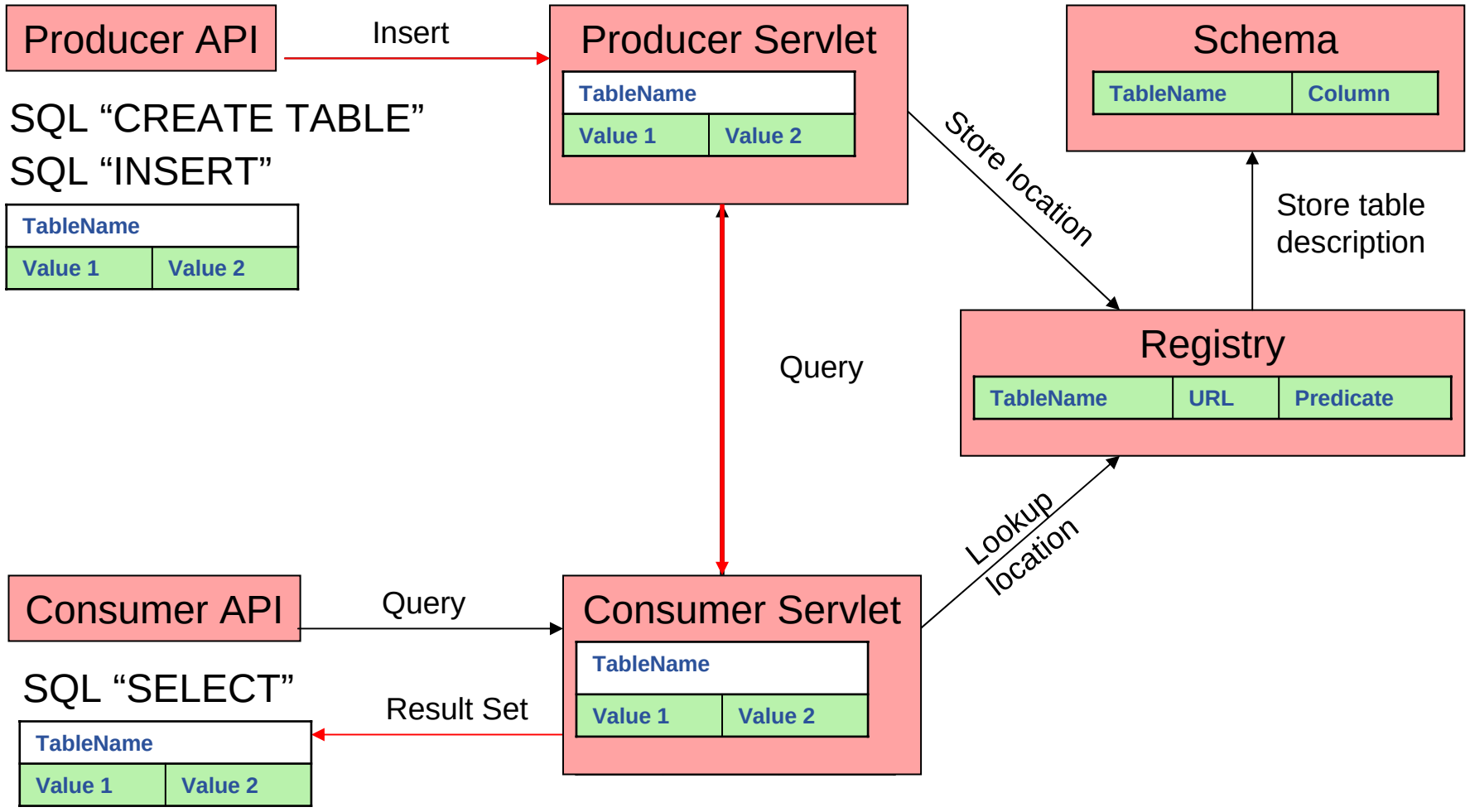

Producer API
Insert
Producer Servlet
TableName
Value 1
Value 2
Schema
TableName
Column
SQL "CREATE TABLE"
SQL "INSERT"
TableName
Value 1
Value 2
Store location
Store table description
Continuous
Registry
TableName
URL
Predicate
Lookup location
Consumer API
Query
Consumer Servlet
TableName
Value 1
Value 2
SQL "SELECT"
Result Set
TableName
Value 1
Value 2

Query Types

- Continuous
- Latest
- History
- Static

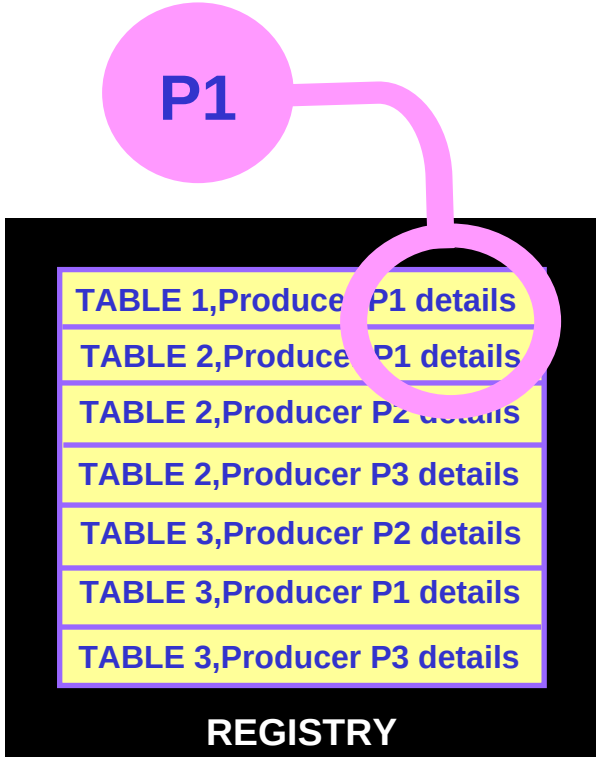


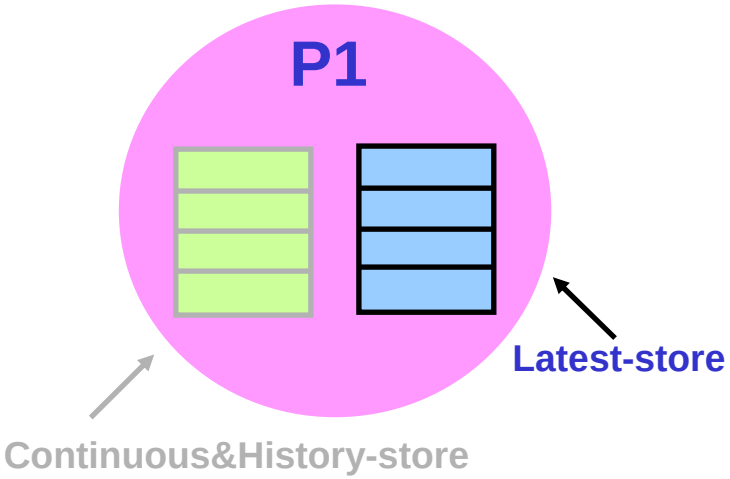
History or Latest



Query Types

- Continuous
- Latest
- History
- Static

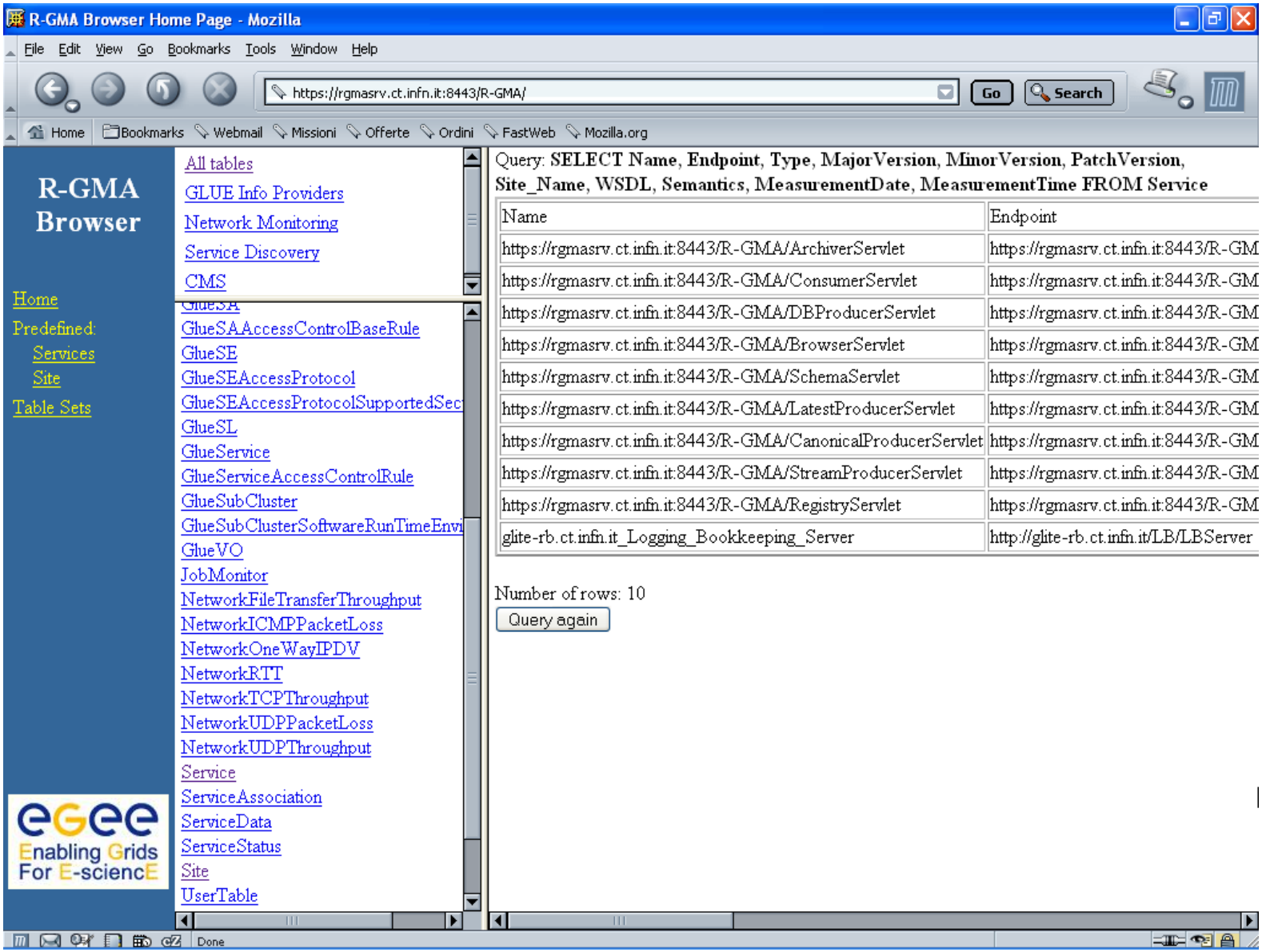




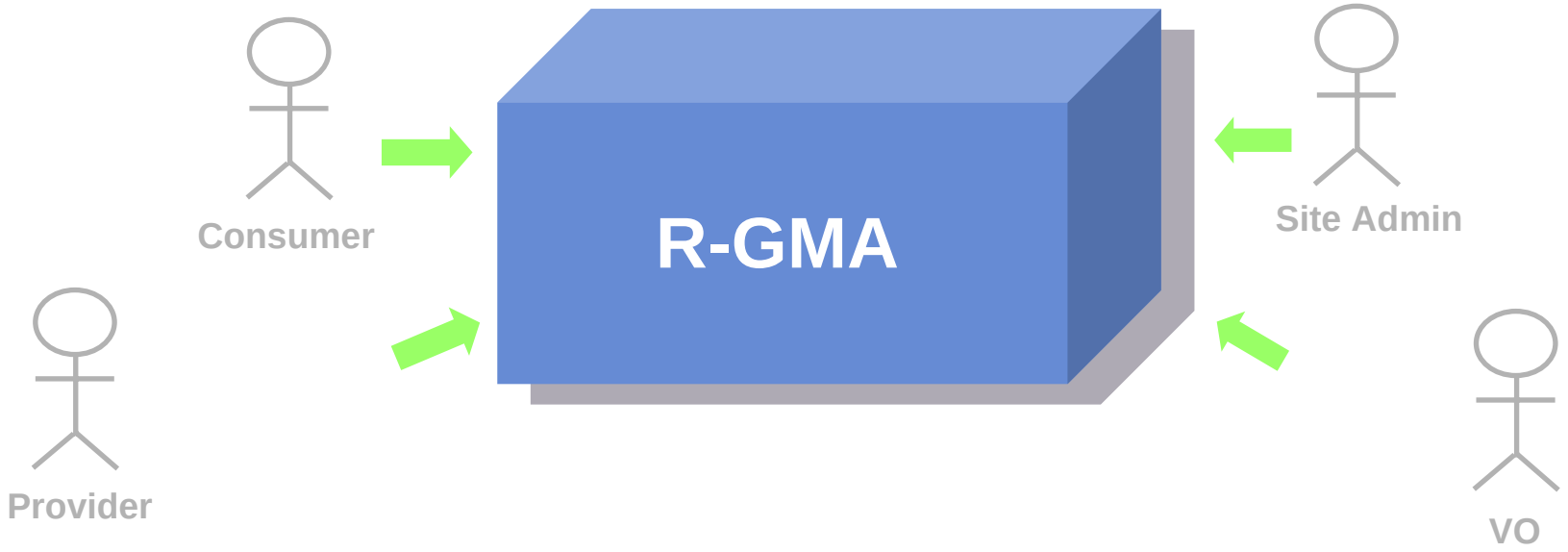
Latest Retention Period

History Retention Period

https://rgmasrv.ct.infn.it:8443/R-GMA

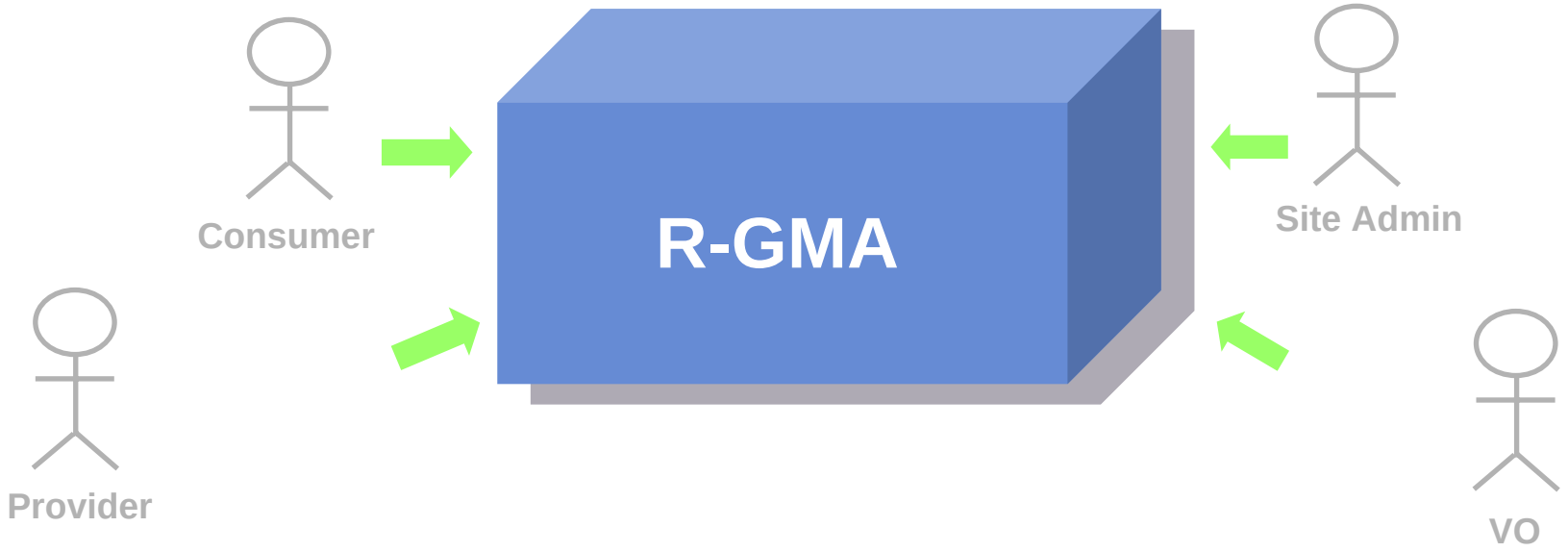


Security: Requirements



- Consumer users: who requests information.
- Producer users: who provides information.
- Site administrators: who runs R-GMA services.
- Virtual Organizations: who “owns” the schema and registry.

Security: Solution



- Mutual Autentication: guaranteeing who is at each end of an exchange of messages.
- Encryption: using an encrypted transport protocol (HTTPS).
- Authorization: implicit or explicit.

More information

- **R-GMA overview page.**
  - http://www.r-gma.org/
- **R-GMA in EGEE**
  - http://hepunx.rl.ac.uk/egee/jra1-uk/
- **R-GMA Documenation**
  - http://hepunx.rl.ac.uk/egee/jra1-uk/glite-r1/