



d4SCIENCE

OGF-Europe Community Outreach Seminar

*Digital Repositories – Interoperability Using
Grid Technologies*

Thursday 5 June, 09.00-17.15, Barcelona, Spain

gCube: A Framework for Controlled Sharing of Data, Services, and Computing and Storage Resources

Pasquale Pagano
CNR-ISTI
Pasquale.pagano@isti.cnr.it



www.d4science.org

- **gCube** was prototyped by the Technical Committee of the DILIGENT project as the **next generation middleware** for supporting **Service Oriented Infrastructure (SOI)**.

SOI is a Virtualised IT Infrastructure which exposes a catalog of services instead of discrete instances, can comprise of SOA Application Support, and includes all configurable infrastructure resources such as compute, storage, and networking hardware and software to support the running of applications.

Consistent with the objectives for SOA, SOI facilitates the reuse and dynamic allocation of necessary infrastructure resources.

- **gCube** is the middleware that powers the **D4Science** infrastructure that will serve EM and FARM VOs.

A Software framework

- to support ON-DEMAND virtual collaborations* among remote parties
 - cost-effective, secure, dynamic, both short and long lived
 - overcome ad-hoc and standard global systems alike
- to exploit an (e-)infrastructure
 - of machines, networks, and application resources
 - promote and/or contribute to resource integration

* Research Environment

SOI

- needs a 'middleware' (typically distributed)
- is open by definition
 - new resource types and/or new resource instances can be de/registered at any time
- is typically layered with respect to functionality
 - control logic, domain services, interfaces

- is powerful if it supports application scope
 - the portion of the infrastructure in which a resource exists
 - the portion of the infrastructure in which a resource can act, operate, or has power or control
- is powerful if it supports sharing scope (controlled resource sharing)
 - connections, machines, storage, data resources, and services

- **gCube** is a **SOI enabling system** to enable the sharing of
 - machines, data, and application services → g3
- **gCube system allows collaborations** in eScience
 - strongly content-oriented, potentially data and processing intensive
- take place in **Virtual Research Environments (VREs)** scope
 - interactively managed, defined, and used:
system administrators, application designers, researchers
 - typically short to medium lifespan
- within the sharing scope of **Virtual Organizations (VOs)**
 - broader and longer lived
 - may stretch across the whole infrastructure
 - or else over significant subsets thereof

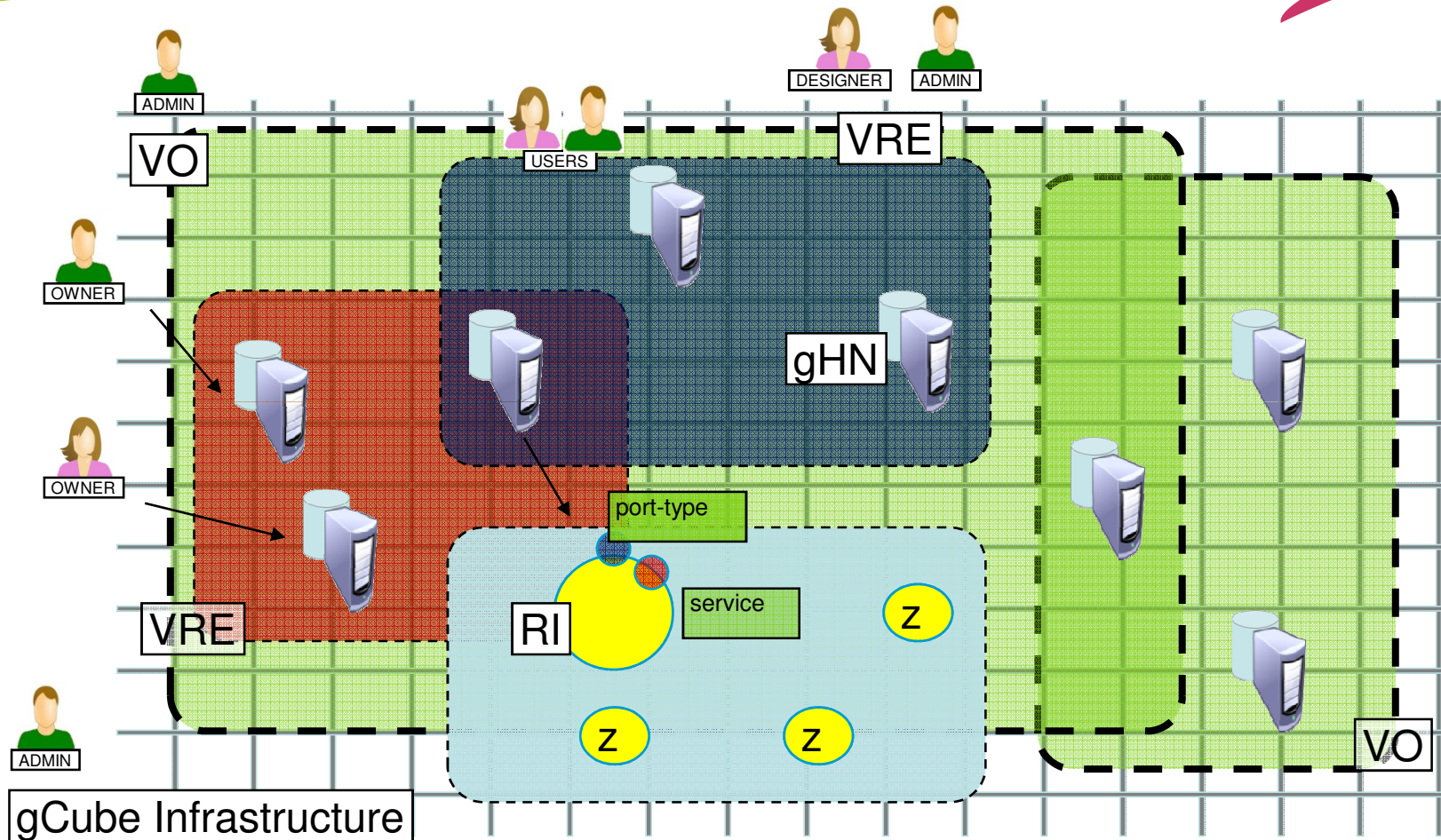
VRE applications are designed as a set of cooperating resources:

- grid resources (computing, storage)
- enabling services (the SOI middleware)
- VRE services
 - content and storage management, search, index, ...
- standalone applications
 - mostly provided by the VOs
- collections of raw data, content, and metadata
 - enriched with schemas, mappings, transformation programs, relationships, ...
- processes defined to manage such resources

- * services are deployed within an infrastructure
 - opportunistically or ubiquitously
 - statically or dynamically

- onto gCube Hosting Nodes (gHNs)
 - runtime containers bound to logical ports
 - provide access to local hardware resources
 - storage, systems, instruments, CPU cycles...
 - instantiate deployed services into Running Instances (RIs)
 - grant basic RI lifetime management
 - mediate RI-to-RI interactions
 - routing requests to target RIs
 - abstract over communication formats and protocols
 - enforce security policies

*: enabling and VRE



- **gCube** creates new opportunities to change the **VRE development model** used by distributed and dynamic organisations and communities
- Using **gCube** empowered infrastructures, the organisations and communities are able to setup their own research environment:
 - When and for the time they need it
 - Exploiting existing Grid-based services
 - Accessing to and handling of distributed multi-focused data and services
 - Orchestrating user defined services, with defined QoS (wrt. scalability, reliability)
 - Profiting from a shared storage and computational set of resources
 - Sharing data and services in a collaborative and efficient way

- gCube is a working horizontal solution
 - composed by 137 software components
 - 60 WSs, 44 libraries, and 33 portlets
 - most components widely tested and certified
 - Public Release (July 2008)
 - providing
 - Service Management
 - Data Management

gCube provides Service Management capabilities

- by enabling:
 - Remote deployment
 - Environment configuration
 - Lifetime management
 - Service provision continuity
- provides concrete solutions for system administrators to
 - eliminate manual deployment overheads,
 - ensure optimal placement of services within the infrastructure
 - support user community services registration and their orchestration by offering a distributed workflow engine
- opens unique opportunities for outsourcing state-of-the-art implementations

SCOPE	ACTION	TIME
Infrastructure	install enabling layer	< 1 day
	install portal	< 1 day
VO	install 1 DHN	< 10 min
	register resources (DHN, data)	< 1 min
	approve resource (DHN, data)	< 1 min
	data publishing (metadata, indexes)	hours
	manage users	< 10 min
VRE	define VRE	< 10 min
	approve VRE	< 1 min
	deploy VRE	< 2 hour
	modify VRE	< 1 hour

- **gCube Data Management** supports researchers by providing means for
 - Persistently storing and physically structuring of content
 - Logical grouping of content in collections
 - Logical sharing of content among several collections
 - Sharing of collections in a VRE and among several VREs through shared workspace
 - Management of complex content consisting of several parts and having multiple representations
 - Storage of structured and heterogeneous metadata compliant with different formats and schemas
 - Programmatic/manual annotation of content via text & images, e.g. data provenance
 - Content linking
 - Definition of 'composite documents' template

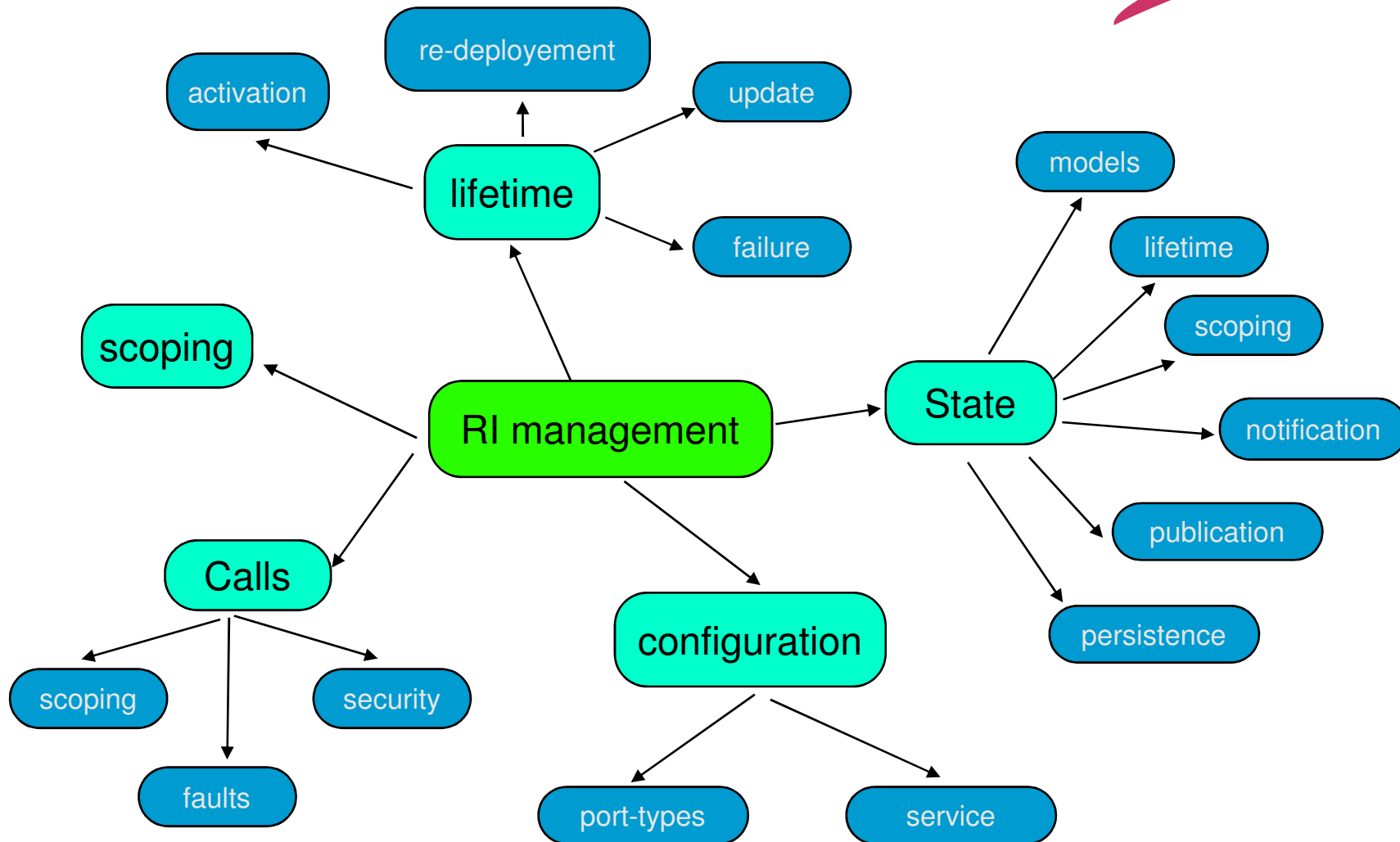
- **gCube Data Management** supports application designers/providers by providing means for
 - Bulk upload and update of data and metadata
 - Manipulation of metadata and data through a powerful Data Transformation Engine (gDTS).
 - Metadata can be cleaned, enriched, and transformed in different formats by exploiting mapping schema, controlled vocabulary, thesauri, and ontology to facilitate data integration and discovery
 - Data can be transformed to offer different views of the same goods
 - Replication and partition of data
 - Subscription and notification
 - Generation and publishing of new data through workflow definition, optimisation, and execution
- gCube exploits Grid technology file-system-like functionalities to manage data storage

An initiative to overcome the complexities of the design and implementation of SOI compliant services

- an application framework for the consolidation / development of existing/new gCube services
 - the **gCube Core Framework (gCF)**

An initiative to meet the needs of system administrators, infrastructure managers, and resource providers

- an easy to install self-contained software distribution for the participation to Service Oriented Infrastructures
 - the **gCube Core Distribution (gCore)**

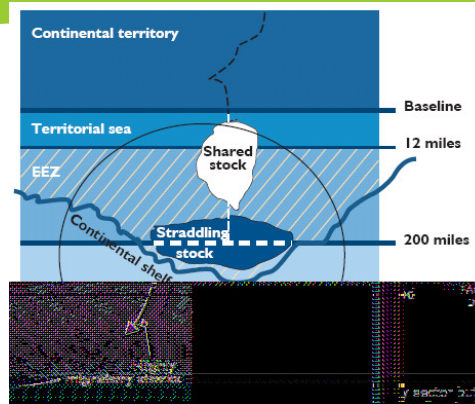


- gCube exploits the Java WSCore, Apache Axis, GridForum specifications (and implementation if any):
 - WS-Notification, WS-Addressing, WS-Security, and WSRF
 - WS-ResourceProperties (WSRF-RP)
 - WS-ResourceLifetime (WSRF-RL)
 - WS-BaseFaults (WSRF-BF)
 - WS-ServiceGroup (WSRF-SG)
 - WS-DAI, WS-DAIR, WS-DAIX
- Mutual authentication based on GSI secure conversation (through delegation and renewal)
- Business Process Execution Language for Web Services (WS-BPEL)
- GridFTP and SRM support
 - VOMS for users and groups management
 - GWT and JSR168(JSR268 is coming)

<https://quality.wiki.d4science.research-infrastructures.eu/quality/index.php/Standards>

- **ISO**: data representation (e.g. ISO3166 for countries, ISO4217 for currencies) and metadata (ISO9115 for GIS)
- **OGF**: Standards related to Architecture (e.g. OGSA), Data (e.g. DAIS, GridFTP), Management (e.g. GLUE, Resources Usage), Applications (e.g. DRMAA), Compute (e.g. JSDL)
- **OAI**: Resources Exposure/Harvesting (OAI-PMH) Resources Exchange (OAI-ORE)
- **OASIS**: Standards related to stateful web services (e.g. WSRF), process management (BPEL), remote user interfaces (WSRP), A&A (SAML / XACML)
- **W3C**: All the standards related to Web Architecture (e.g., URI/URL, HTTP), Service Oriented Architectures (e.g. SOAP, WSDL, WS-Addressing) and data representation and manipulation(e.g. XML*)
- **Others**: Classification systems (e.g. ISSCAAP, ISSCFV, ISSCFG), features representation (e.g. GML for GIS), metadata (e.g. AgMES for Agricultural, SDMX for Statistics)

- **gCube** reduces the costs to manage any complex multi-domain Service Oriented Infrastructure.
- **gCube** offers an horizontal solution to manage and enrich on-demand created VREs on complex Service Oriented Infrastructures.
- **gCube** is equipped with data and metadata management facilities that allows to make interoperable heterogeneous data sources.
- **gCube** is compliant with consolidated and emerging standards.
- **gCube** offers an open family of frameworks that can be easily customised



www.gcube-system.org

