



d4SCIENCE

**DL.ORG
Functionality Group
Meeting**

29-30 June 2009
Athens (GR)

**gCube:
Functionality & Interoperability**

George Kakaletris
NKUA
gkakas@di.uoa.gr



- D4Science project & concepts
- gCube at a glance
- gCube Functionality Overview
- Interoperability support in gCube
- From D4Science to D4Science II : What is next ?

Sequel of **DILIGENT** FP6 IST Project

EGEE associated project

Funding: EC/FP7/ICT/eInfrastructures

Duration: 24 Months (1/1/2008–31/12/2009)

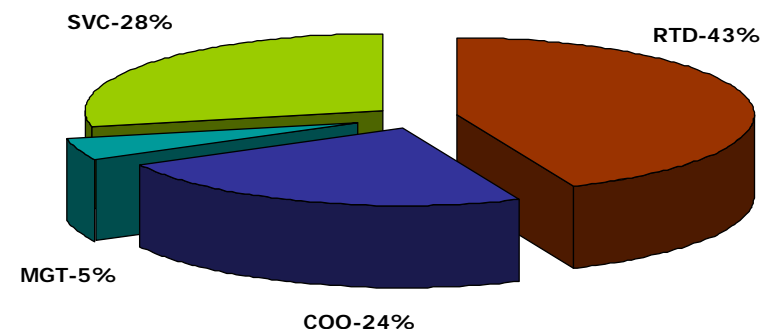
Effort: 400MM

- Total cost: ~3.765 K€
- Maximum EC funding: 3.150 K€

11 Partners

- Management (ERCIM)
- RTD (CNR-ISTI, Univ. Athens, Univ. Strathclyde, Engineering, Univ. Basel)
- Services (CERN, 4D Soft)
- User Communities (ESA, FAO, WorldFish)

Full title: **D**istributed
collaboratories
Infrastructure on
Grid **EN**abled
Technology *for*
Science



- Consolidate & extend **gCube** software produced by DILIGENT to reach production-level quality
 - Release the software for internal and external use
- Analyze and implement three distinct application scenarios in two different domains:
 - Environmental Monitoring
 - Fishery Resources Management
- Deploy the infrastructure and serve the named application scenarios

An eInfrastructure for enabling the sharing of existing and creation of new knowledge among its served scientific community members.

- Potential characteristics of VREs:
 - Hosting on distributed eInfrastructures
 - Spanning multiple physical organisations / groups
 - Employment of domain-specific tools and applications
- Common patterns of requirements in VREs:
 - “Resources” and their control
 - Persistence/access of/to data, information, knowledge
 - ...

gCube

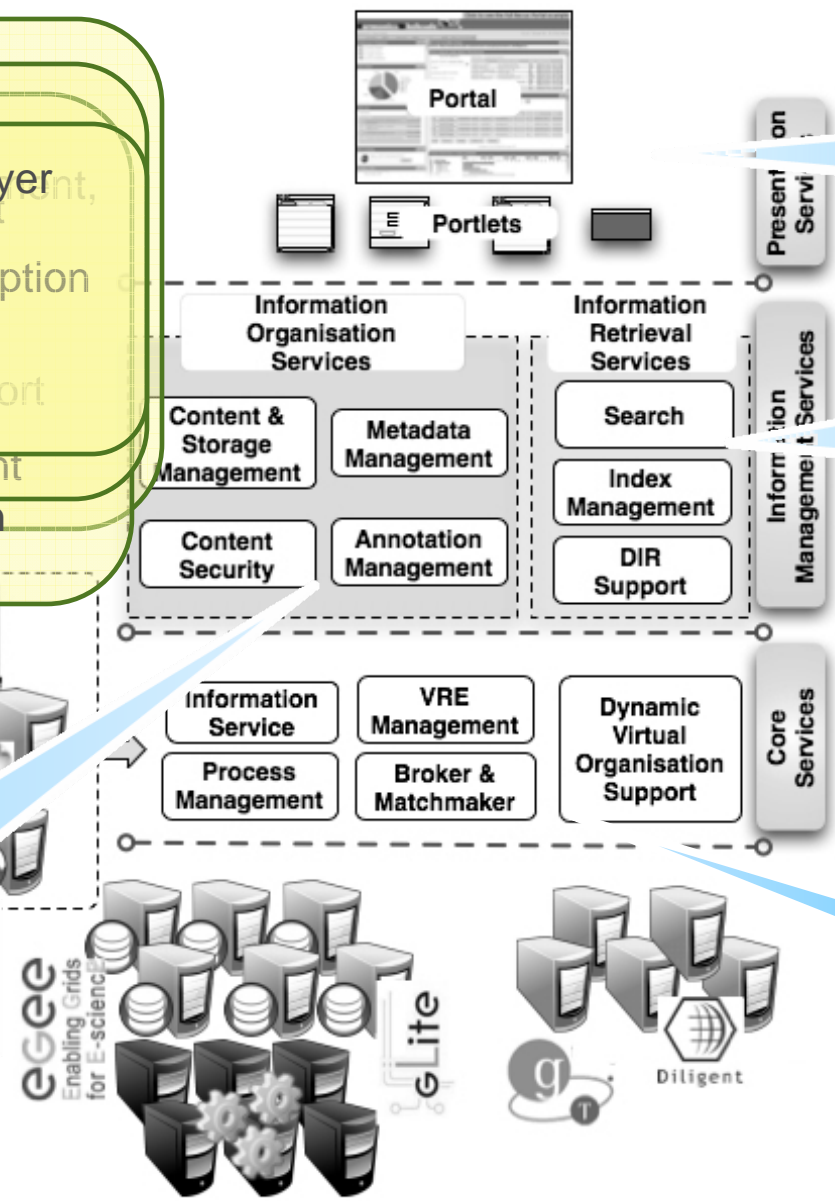
a system to enable **e-infrastructures** for the creation, hosting and maintenance of dynamic **virtual environments** capable of satisfying complex research and collaboration needs of distributed **scientific communities** organized in **Virtual Organizations** (VOs)

Architectural & technological background

- Service Oriented Infrastructure
- The Grid



- ✓ Runtime Environment
- ✓ Metadata Indexing
- ✓ Application Support Layer
- ✓ Portals
- ✓ User Portlets
- ✓ Administrative Portlets
- ✓ Desktop clients
- ✓ Search
- ✓ Annotation Management
- ✓ Content Transformation
- ✓ Ontology Management



Presentation Services

Information Retrieval Services

Information Organisation Services

Core Services

- Hosting “data” collections
 - Import (Linking/embedding, partial/full, updating)
 - “Preservation”
 - Evolution / Curation (incl notifications)
 - Browsing / Retrieval (next topic)
 - Delivery
- Challenge: Nature of “data”:
 - Metadata can be of any manifestation
 - Even “unstructured”
 - Content can be any nature (yet always digital 😊)
 - Even raw data
 - Sources communication can employ any protocol
- The enabling parts:
 - Information Object model (hosted in Content Management Service)
 - Highly versatile Archive Import Service (plugins / scripting based)
 - Format agnostic Data Processing Pipeline (DPP), Transformation & Presentation Facilities

- Locate any information stored in the gCube Information Space & affiliated sources
 - Fully/semi qualified searches on metadata
 - GeoSpatial / temporal search
 - “Similarity” search on content (CBIR)
 - Extract-based (full text search)
 - (Multilingual) NLP
 - Feature based (other content)
 - Federated Search / Fusion of results
- Challenges
 - No common schema is enforced by the system or “feasible”
 - No common presentation needs
 - Binary content can be literally everything
 - No common languages & capacities among “sources”
- The approach: “Operator”-based search
 - Powered by the gCube DPP

- Maintain personal workspaces to power VRE exploitation and collaboration
 - Place Information Objects in the workspace
 - Search in the workspace contents
 - Share items in the workspace among users
- Enabled by:
 - Workspace management service
 - Workspace frontend (portlet)
 - Persistent ResultSets
 - Peripheral facilities (Content viewers, annotators, ...)

- Annotate any content stored in the gCube Information Space
 - Anchored & non anchored annotations
 - Annotate via text / graphic elements / other objects / ontologies
 - Protected & shared annotations
- Challenges:
 - Arbitrary content in gCube Information Space: no single way to present and anchor annotations
 - Video, Images (any type), DataSets, Rich Documents (PDF, DOC etc), Sound ...
- Partial achievement via:
 - Annotations managed as Metadata
 - gCube Annotation Backend / FrontEnd (ABE & AFE)
 - Supporting Services (Transformation, Ontology Management)
 - *No protection feasible currently*

- Exploit the data of the VRE to build (periodic) scientific reports
 - Report templates
 - Notifications / IR / Custom processing for populating parts of reports
 - Modular report nature (different assignees)
 - Protected sections
 - Manual editing of final product
 - Sharing of report product
- Functionality is currently underway
 - XML Report Templates
 - gCube DPP for retrieving parts

- Curation of timeseries
 - Ontologies
 - Reference data
- MGVI Processing
 - Satellite data
- Acuamaps production
 - Statistical curated catch data
- Custom workflows

- On-demand created VREs
 - Selection of resources and members
 - User friendly procedures for non core-technical personnel
 - Substantial differentiation among VREs (presentation, services, resources)
- Solution:
 - Hierarchical infrastructure: *Infrastructure* → *VO* → *VRE*
 - VREs as subsets of VOs
- Capacities
 - Scoping of VO level resources (security / visibility)
 - Customisation of resources granted at the VO Level
 - Automatic demolition (on date)

- VO / VRE Management facilities
 - Editing / Creating / Assigning resources
 - Adding / Removing / Enabling users
- System behavior
 - Automatic reconfiguration (zero-administration needs)
 - Optimal resource use
- On-demand content transformation
- Ad-hoc processing during Information Retrieval process
- Personalisation
- Integration of diverse content stores (Grid Storage Elements, FTP sites, ...)
- eLearning Integration
- Content Protection

Integratability ● Interoperability ● Composability

Two orthogonal fields of action in D4Science:

- “Data”
 - Information Object Model
 - Manifestations of Metadata/Content/Data payload
 - Resource Model
- “Logic”
 - Resource Model
 - Communication protocols
 - Software components technology

- The foundations:
 - Flexible (yet rather “flat”) Information Object Model
 - Data Agnostic Core-System
 - “Standards”
- Major affected areas
 - Data import & hosting
 - Data Processing Pipeline (aka Search Engine)
 - Data delivery & presentation
- Beyond the agnostic system:
 - Specialization by “pluggable” logic ...

- The foundations
 - “Standards”
 - Open resource model
 - Incl. “Semantic” information
 - Plugability principles
- Major affected areas
 - System Architecture (WSRF-based SOA)
 - Multiple APIs offered
 - Enabling services
 - All higher services (exploitation)

- WS-*
- WSRF
- X-*
- WS-BPEL
- JSR
- Glue Schema
- GSI-Security
- OpenSearch

- Java
- Globus Toolkit
- gLite

- Distributed under Open Source License
 - EUPL



More Exploited:

- DC
- ISO

Under development

- OAI-PMH

Under consideration

- OAI-ORE
- WS-DAI
- OpenGIS - related

Quick Info

- Duration: 24 months
- Total budget: 5 089 700 €
- EC contribution: 4 300 000 €
- Time plan: Oct 2009 – Sep 2011

Major objective:

“To advance D4Science e-Infrastructure into a pivotal element that, through its capabilities and mediating role, will be capable of drawing numerous infrastructural initiatives and scientific communities within the scope of a **Knowledge Ecosystem**” (D4Science II DoW)

that is:

gCube in D4Science II will enable interoperability among current (D4Science, DRIVER, GENESI-DR, INSPIRE, AquaMaps) and future eInfrastructures for the cross fertilization of scientific data and approaches.

Thank you