



# **D4Science: a Data Infrastructure Ecosystem for Science**

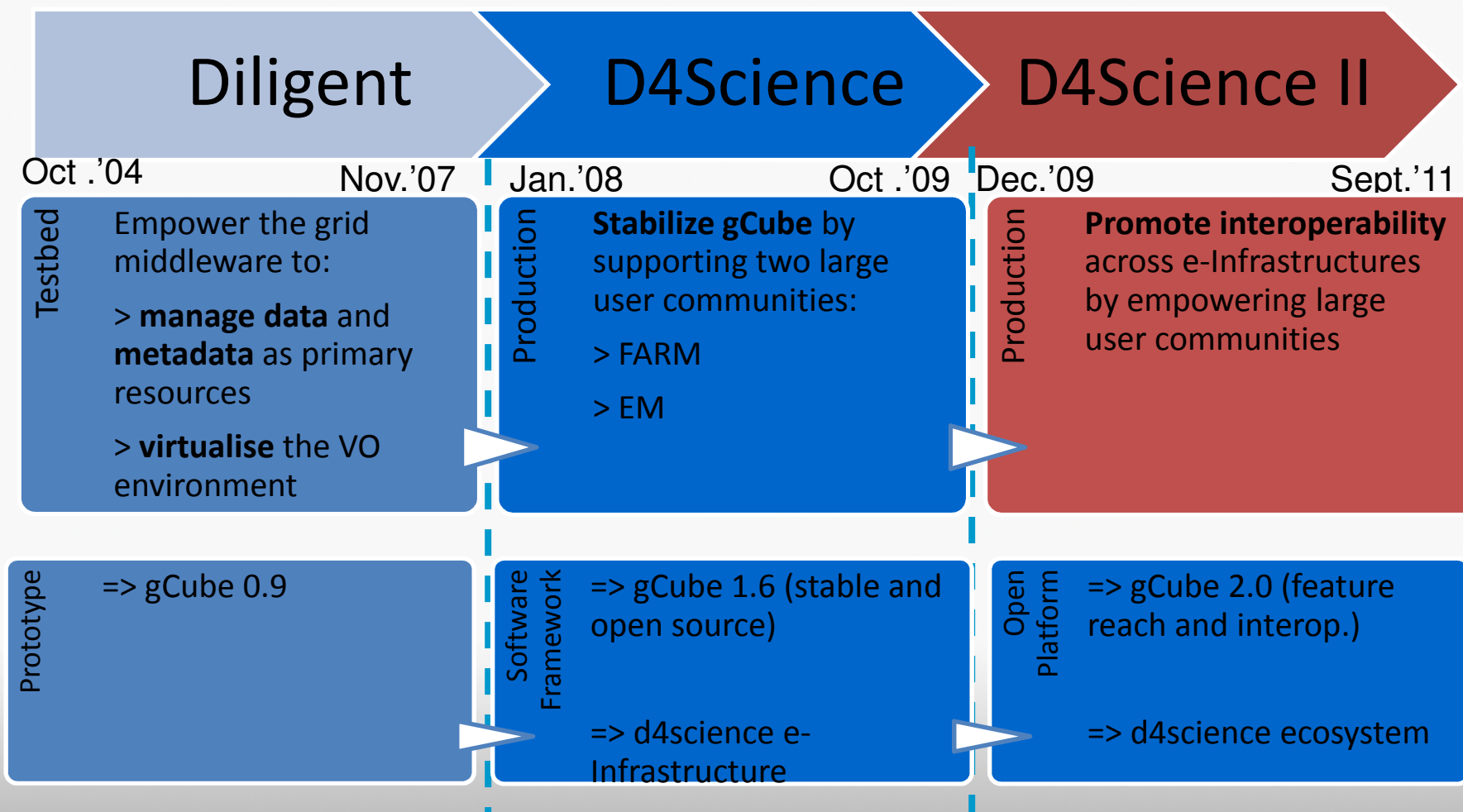
**Leonardo Candela**

**6<sup>th</sup> October 2010**

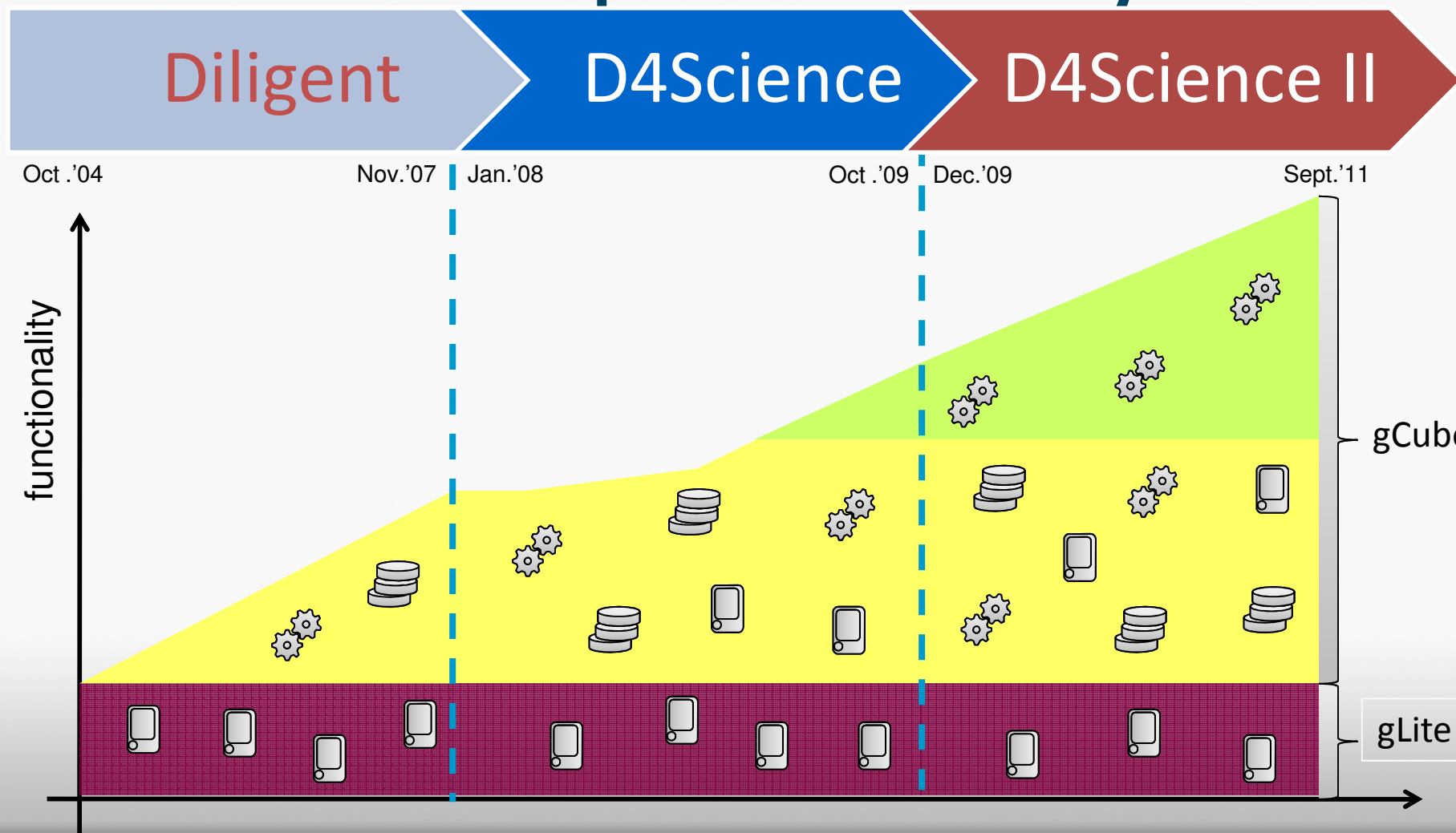
**DL.org Autumn School – Athens, 3-8 October 2010**

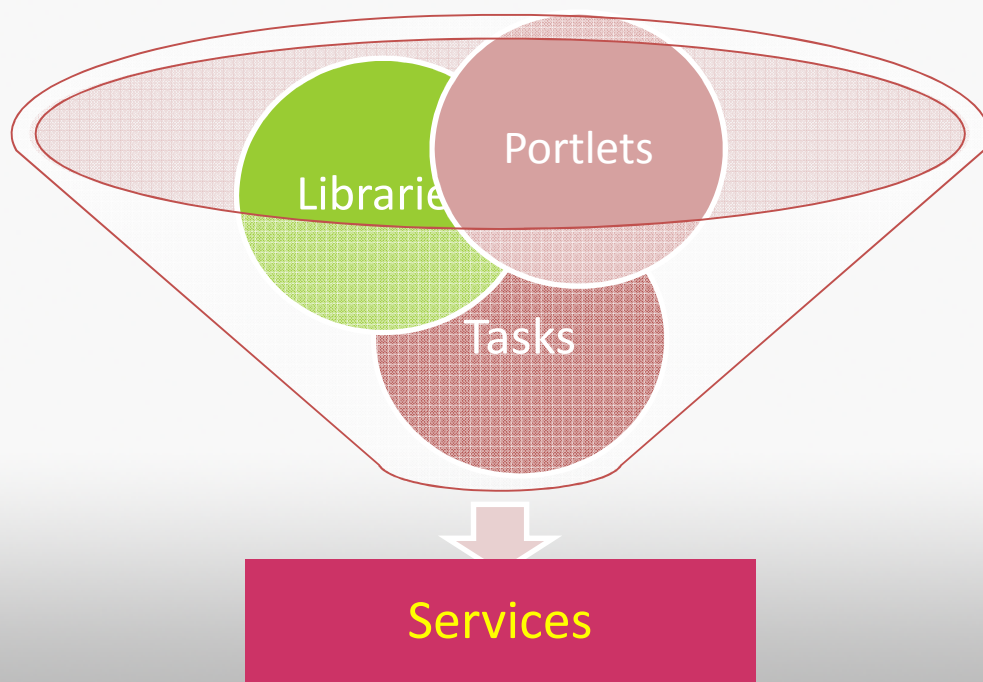


# From a testbed to a production ecosystem



# From a testbed to a production ecosystem





- 395 software components
- In production
- Multiple communities

## Infrastructure vs. gCube e-Infrastructure

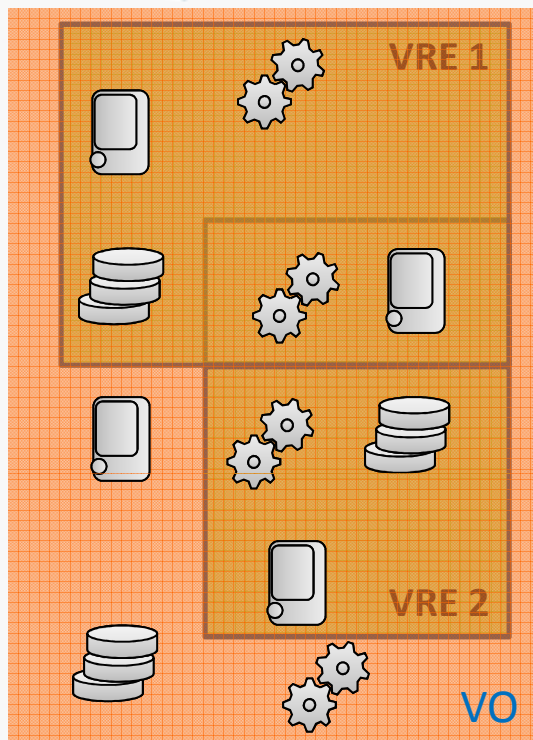
- An infrastructure is the basic physical and organizational structures and facilities needed for the operation of an organization.
- A gCube e-Infrastructure promotes effective consumption of shared resources:
  - hardware resources
  - data resources
  - software resources

**gCube promotes its own services to shareable resources on par with data and the hardware resources**

# Serving Communities through Virtual Research Environment

Is a tailored and consolidated environment adequate to represent a growing aggregation of resources tailored to satisfy the **evolving needs** of the user community?

Virtual Research Environments



# Virtual Research Environment

Virtual Research Environment (VRE) is

- a distributed and dynamically created environment
- where subset of resources can be assigned to a subset of users via interfaces
- for a limited timeframe
- at little or no cost for the providers of the infrastructure

**gCube is a first example of a VRE Management System**



# gCube Completeness: A full view of gCube

## Presentation Services

- ✓ Application Support Layer
- ✓ User Portlets
- ✓ Administrative Portlets
- ✓ Desktop clients

## Information Retrieval Services

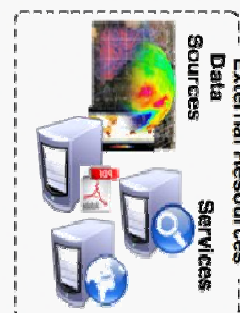
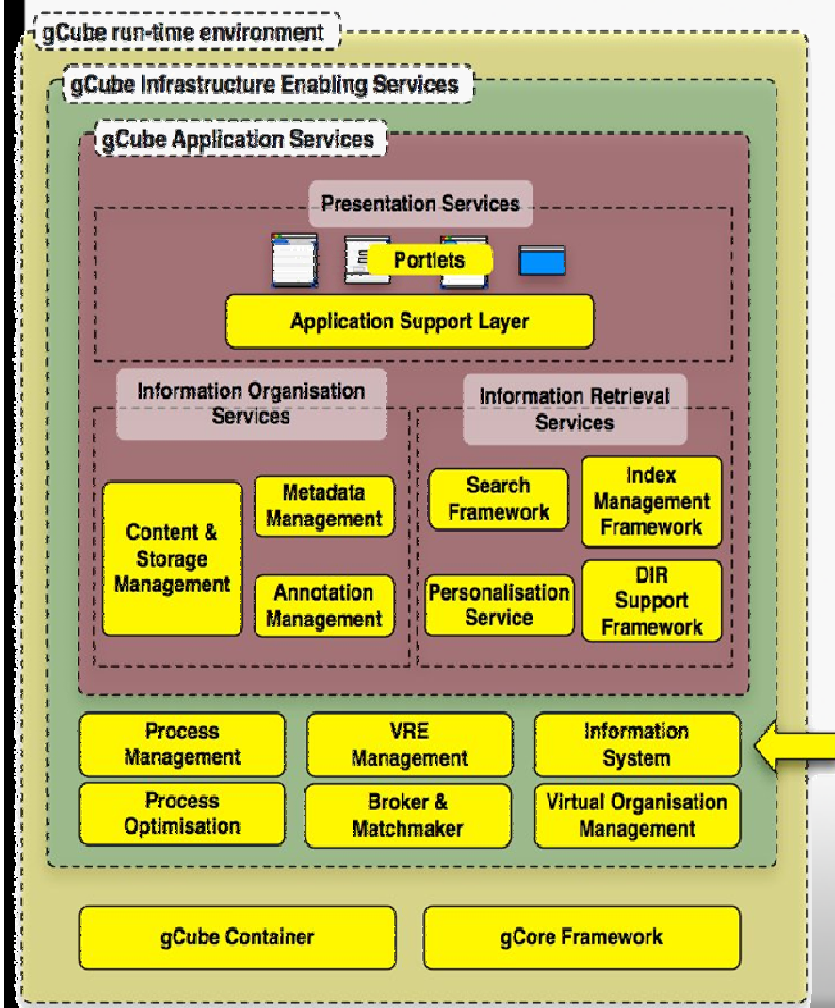
- ✓ Metadata Indexing
- ✓ Content Indexing
- ✓ Personalisation
- ✓ Content Source Description & Selection
- ✓ Data Fusion
- ✓ Search

## Information Organisation Services

- ✓ Storage Management
- ✓ Collection Management
- ✓ Content Management
- ✓ Metadata Management
- ✓ Archive Import
- ✓ Metadata Brokerage
- ✓ Annotation Management
- ✓ Content Transformation
- ✓ Ontology Management

## Enabling Elements

- ✓ Runtime Environment provision (gCore/gHN)
- ✓ Infrastructure Management, Monitoring and Self-reorganisation
- ✓ VRE Management
- ✓ VO and Security Support Services
- ✓ Process Execution

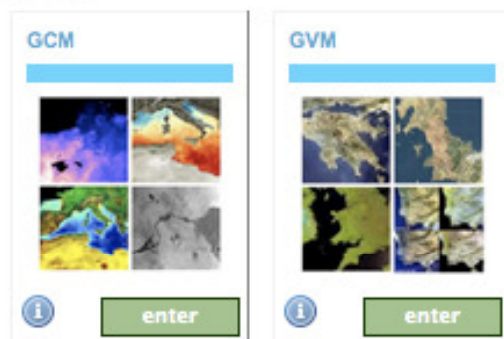




# Infrastructure Exploitation

## Production

EM 



FARM 



### Nodes



- 30 Nodes
- CNR
  - NKUA
  - ESA
  - FAO
  - UNIBASEL

- 29 Nodes
- CNR
  - NKUA
  - FAO
  - UNIBASEL

### Collections



- 25 Data
- EEA
  - MERIS
  - AATSR
- 69 Metadata
- es
  - ISO19115
  - eiDB

- 15 Data
- AquaMaps
  - Fact Sheets
  - Country Maps
- 28 Metadata
- FARM\_dc
  - aquamaps

### Functionality



- Integration with gPod
- Geographical and text search
- Search by metadata
- Personal workspace
- Objects annotation
- Report generation
- Maps Generation

**More than 500 Web Services  
autonomically managed**

## gCube is ...

Distributed Software System

*Large scale service-based*



Infrastructure Enabler

*Resource-Rich*



VRE Management System

*Autonomic, Extensible, and Maintainable*

# Building Virtual Research Environments

The candidate GHN for VRE Manager deployment is node13.p.d4science.research-infrastructures.eu:

## ☐ Hostname

< 11 - 19 of  
prev 19

- |   |                         |
|---|-------------------------|
| <input type="checkbox"/> dl27.di.uoa.gr:8084  | <a href="#">Actions</a> |
| <input checked="" type="checkbox"/> node13.p.d4science.research-infrastructures.eu:8080 | <a href="#">Actions</a> |
| <input type="checkbox"/> node22.p.d4science.research-infrastructures.eu:8080            | <a href="#">Actions</a> |
| <input type="checkbox"/> node25.p.d4science.research-infrastructures.eu:8080            | <a href="#">Actions</a> |
| <input type="checkbox"/> node29.p.d4science.research-infrastructures.eu:8080            | <a href="#">Actions</a> |
| <input type="checkbox"/> node31.p.d4science.research-infrastructures.eu:8080            | <a href="#">Actions</a> |
| <input checked="" type="checkbox"/> node8.p.d4science.research-infrastructures.eu:8080  | <a href="#">Actions</a> |
| <input type="checkbox"/> portal.d4science.research-infrastructures.eu:9000              | <a href="#">Actions</a> |
| <input type="checkbox"/> wn02.research-infrastructures.eu:9001                          | <a href="#">Actions</a> |

## node22.p.d4science.research-infra

**Security:** false  
**Up Time:** 112 days

### Memory

**Virtual Available:** 1372 KB  
**Virtual Size:** 1488 KB

### Site

**Location:** Pisa  
**Country:** it  
**Domain:** p.d4science.research-infrast

### RunningInstances

Start

VRE Information

VRE Content

-VRE Collections

-VRE Metadata  
Formats

VRE Functionality

VRE Architecture

Finish

Cancel

<< Back

Next >>

Create

# Cooperative Tools in a Nutshell

A virtual desktop to organize the working environment

VRE UI

Workspace

Tools supporting generic as well as specific tasks

Search

Annotation

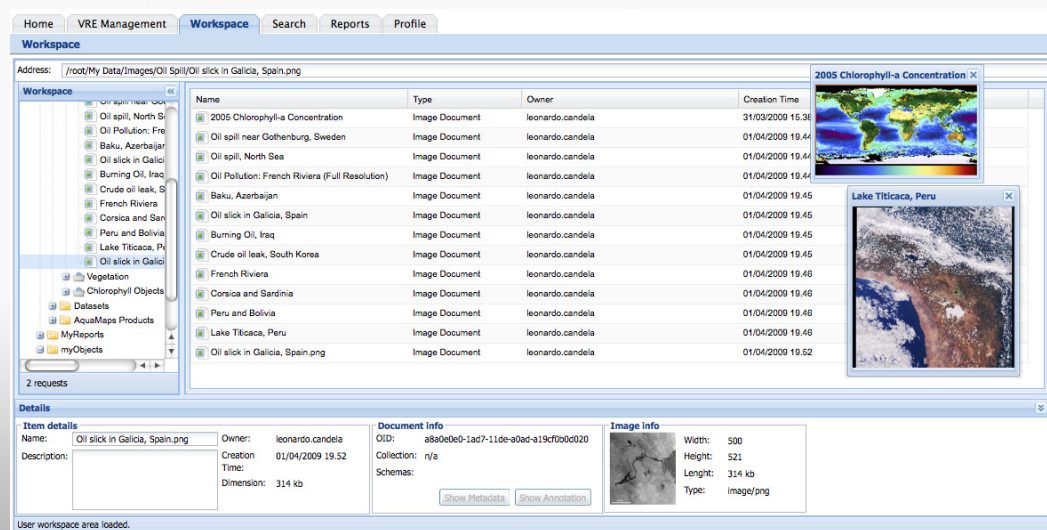
Biodiversity  
Map

A virtual document to innovatively  
describe research packages

Report  
Management

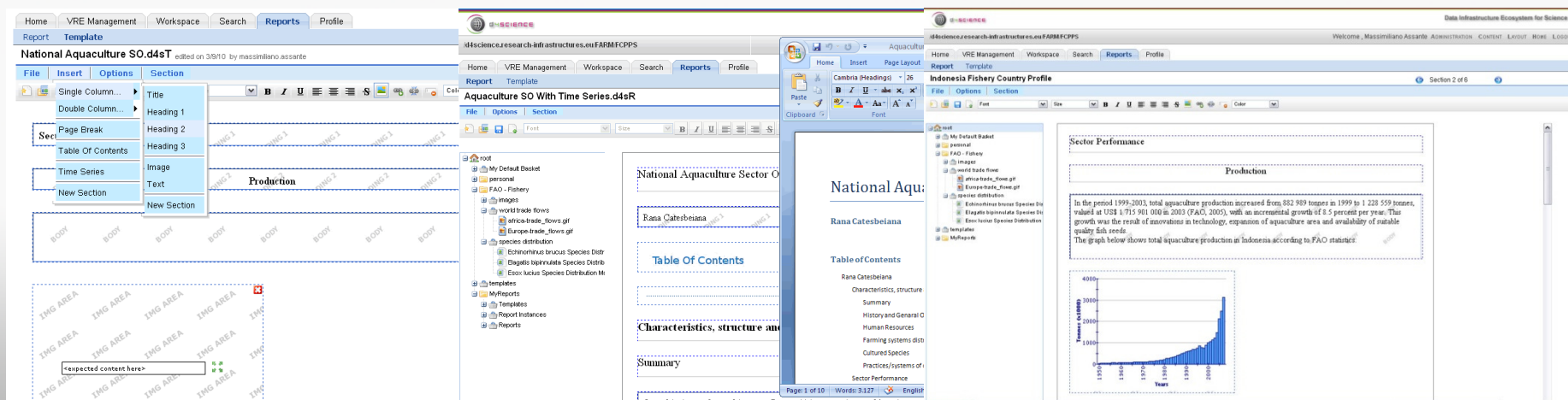
# Workspace: the service

- A collaboration-oriented suite providing for
  - seamless access and organisation facilities on a rich array of objects (e.g. Information Objects, Queries, Files, Templates)
  - mediation between external world objects, systems and infrastructures (import/export/publishing)
  - support common file manager (drag & drop, contextual menu)
  - support an effective rich object sharing facility



# Report Management: the service

- A collaboration-oriented suite providing for
  - template-oriented, feature-rich and flexible document format definition
  - effective and infrastructure-integrated report compilation (drag & drop workspace items)
  - collaborative and distributed editing (workspace based)
  - standard-based report materialisation (HTML, OpenXML)



The screenshots illustrate the DL.org Report Management service interface. The left panel shows a template editor for 'National Aquaculture SO.d4sT' with a sidebar menu for sections like Title, Heading 1, Heading 2, Heading 3, Text, Image, and Table Of Contents. The middle panel shows a workspace with a file tree on the left containing folders like 'personal', 'FAO - Fishery', and 'world trade flows', and a central preview area showing a 'National Aquaculture Sector Overview' report. The right panel shows a 'Indonesia Fishery Country Profile' report with a 'Sector Performance' section, a 'Production' subsection, and a bar chart showing production from 1999 to 2003.

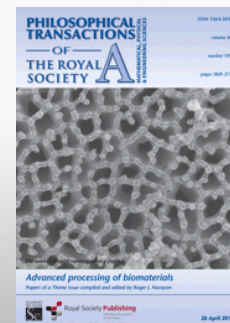


# VREs, Workspaces and Report in Action



# gCube and Humanities: the gMan case

- JISC - King's College London
- Look at new ways of integrating existing data resources for Classics and add services so that research work based on integrated resources can be published
- Data sources
  - The Heidelberger Gesamtverzeichnis (HGV) der griechischen Papyrusurkunden Aegyptens, a collection of metadata records for 55,000 Greek papyri from Egypt.
  - Projet Volterra, a database of Roman legal texts, and associated metadata, from various sources (epigraphic, papyrological, or literary) currently in the low tens of thousands but very much in progress.
  - The Inscriptions of Aphrodisias, (InsAph), a corpus of about 2,000 ancient Greek inscriptions from the Roman city of Aphrodisias in Asia Minor, including transcribed texts and metadata marked up using EpiDoc TEI, as well as images of the physical objects.
- Main functionality
  - cross-collection search
  - workspace
  - annotation
  - report creation
- Early results in “AHM 2009 Phil. Trans. A special issue”



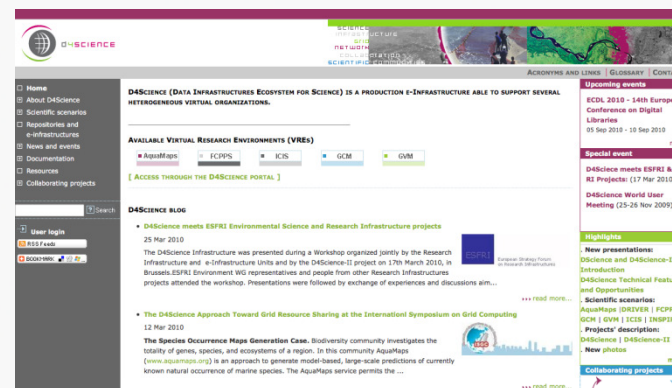


# gCube / D4Science

[www.gcube-system.org](http://www.gcube-system.org)



[www.d4science.eu](http://www.d4science.eu)



Donatella Castelli  
D4Science-II Project Director  
[donatella.castelli@isti.cnr.it](mailto:donatella.castelli@isti.cnr.it)

Pasquale Pagano  
D4Science-II Technical Director  
[pasquale.pagano@isti.cnr.it](mailto:pasquale.pagano@isti.cnr.it)