DL.org

Digital Library Interoperability,
Best Practices and Modelling Foundations

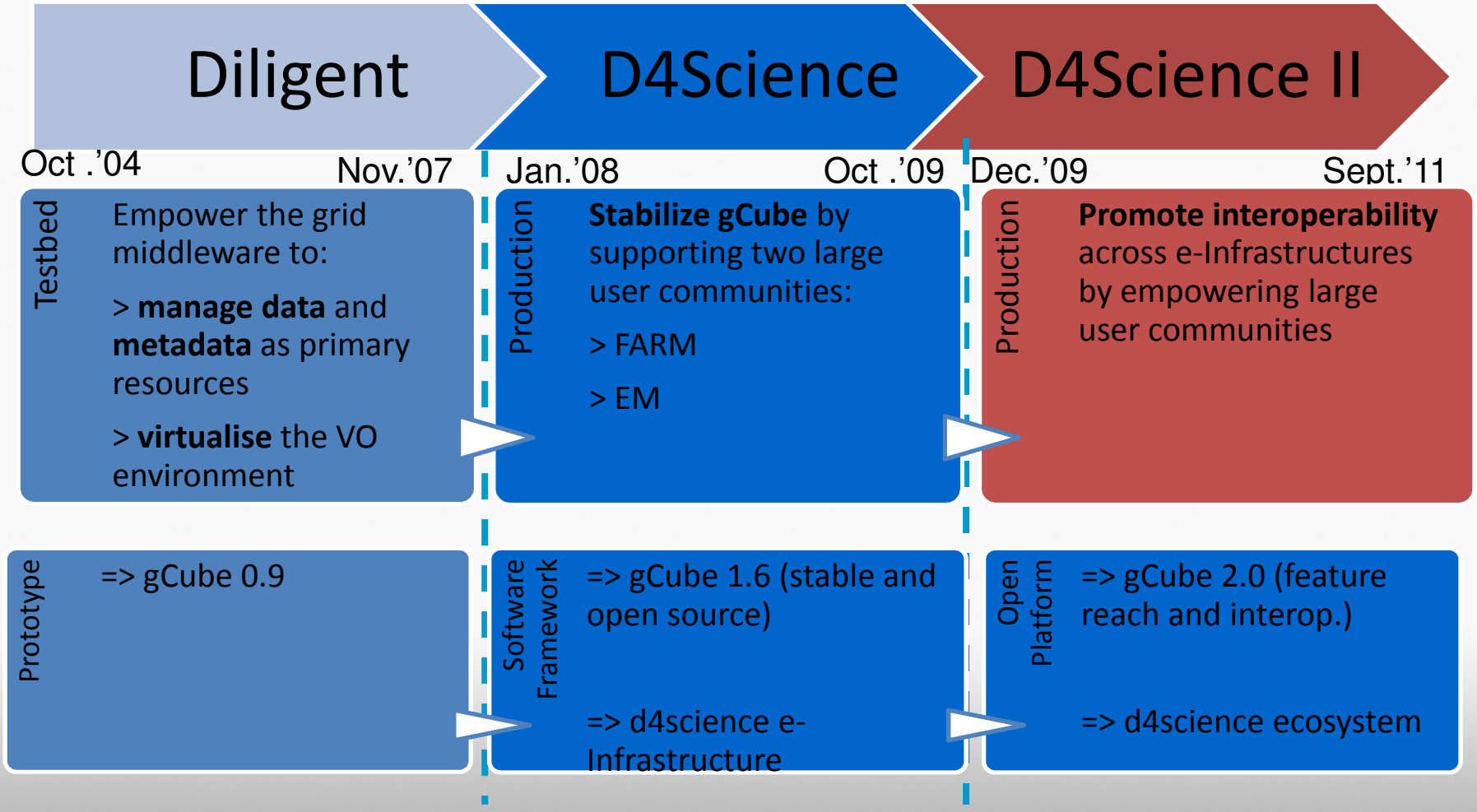
D4Science: a Data Infrastructure Ecosystem for Science

Leonardo Candela

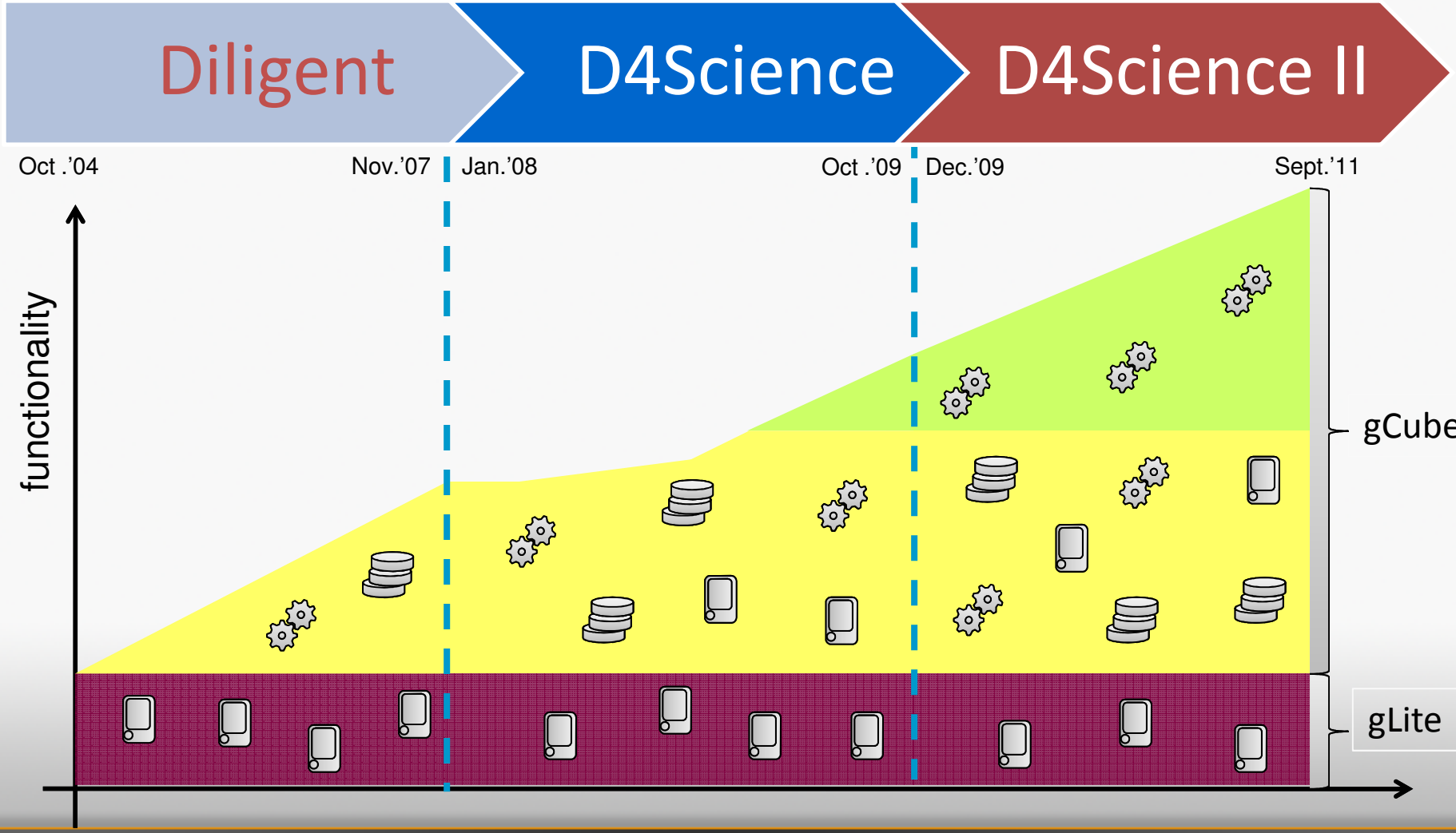
6th October 2010

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

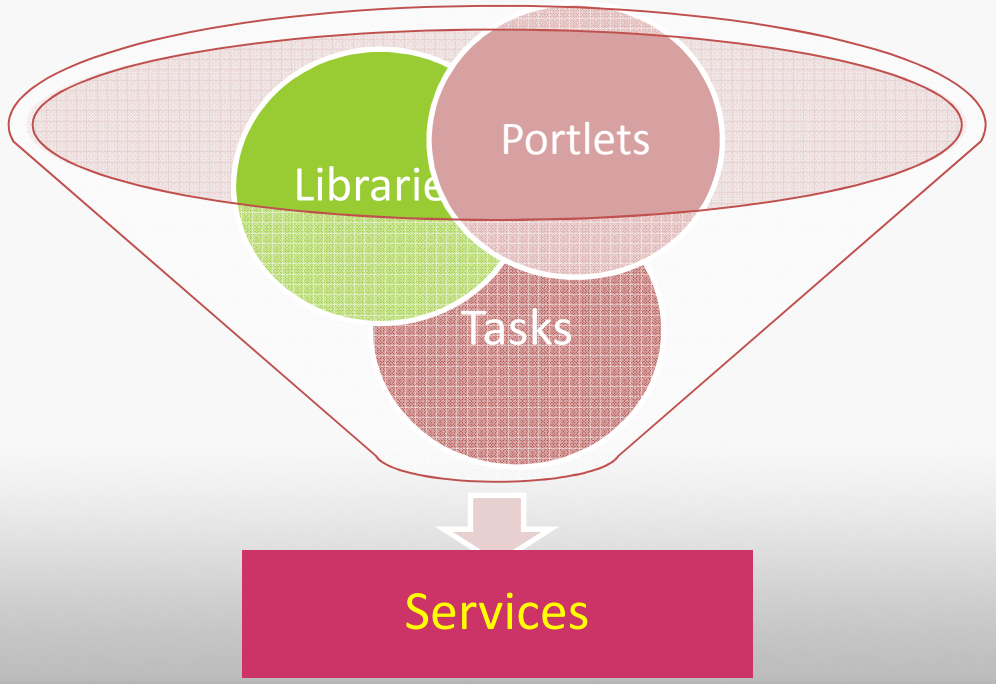
From a testbed to a production ecosystem



From a testbed to a production ecosystem



www.gcube-system.org



- 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

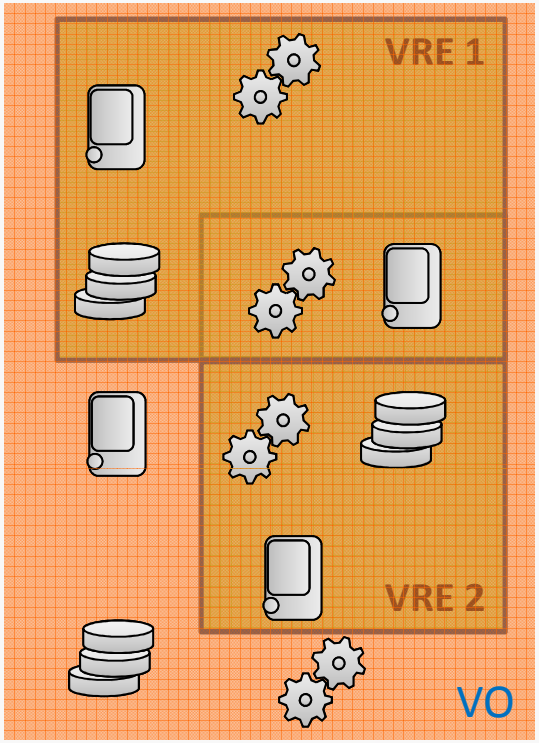
gCube promotes its own services to shareable resources on par with data and the hardware resources

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

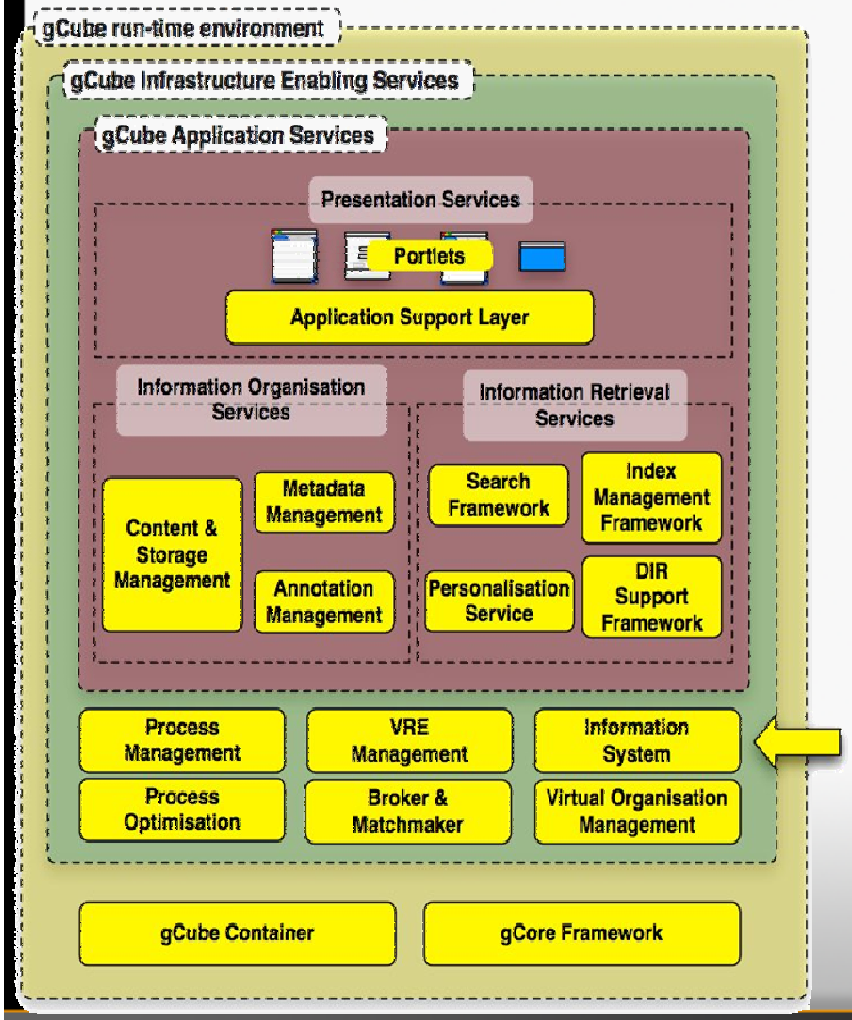


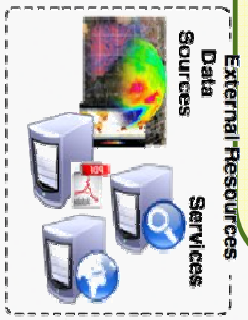
Virtual Research Environment (VRE) is

- 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

Nodes	Collections	Functionality
30 Nodes • CNR • NKUA • ESA • FAO • UNIBASEL	25 Data • EEA • MERIS • AATSR 69 Metadata • es • ISO19115 • eiDB	• Integration with gPod • Geographical and text search • Search by metadata • Personal workspace • Objects annotation • Report generation
29 Nodes • CNR • NKUA • FAO • UNIBASEL	15 Data • AquaMaps • Fact Sheets • Country Maps 28 Metadata • FARM_dc • aquamaps	• Maps Generation

More than 500 Web Services autonomically managed

DL.org
Digital Library Interoperability, Best Practices and Modelling Foundations

gCube is …

Distributed Software System

Large scale service-based

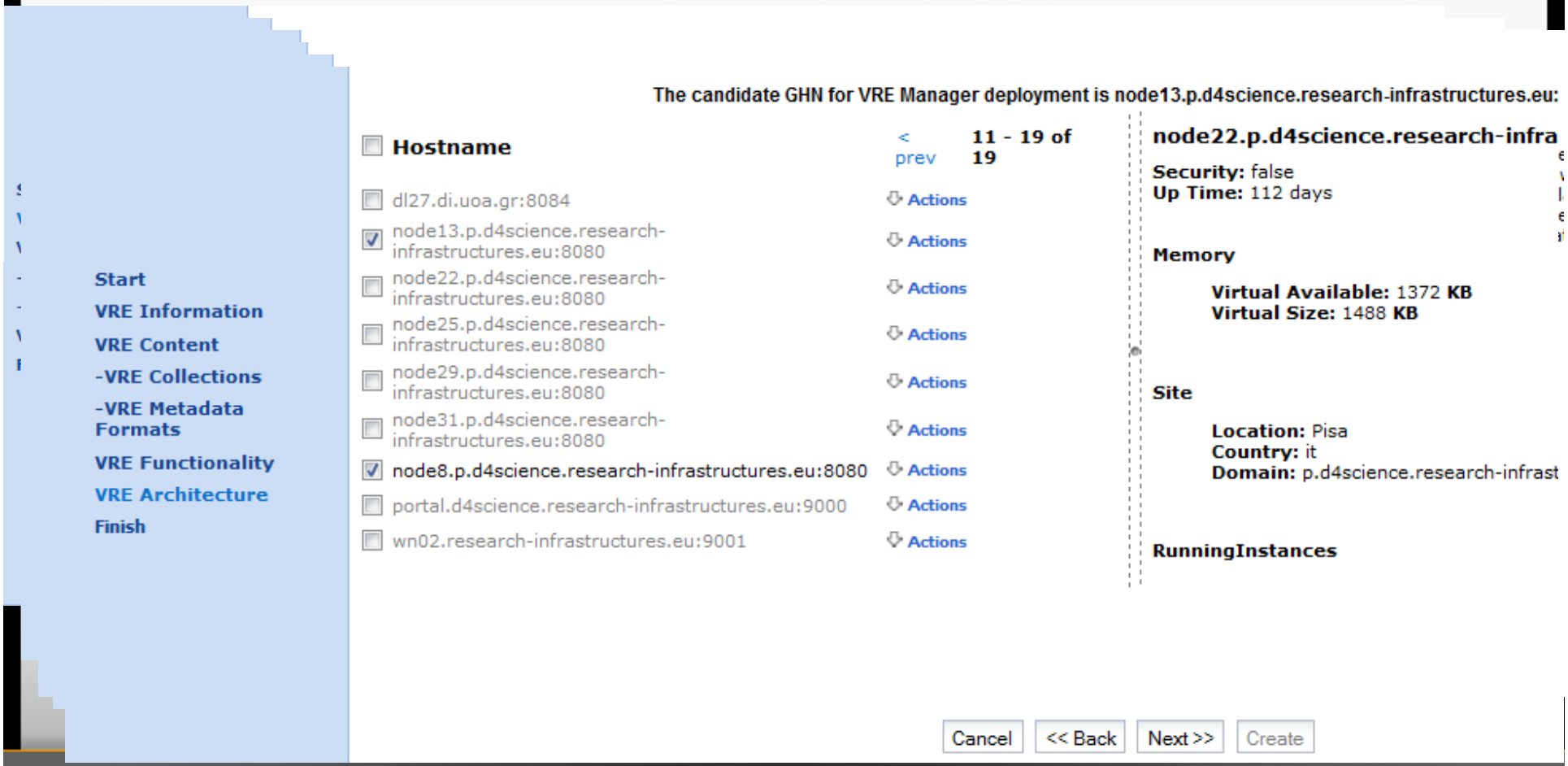
Infrastructure Enabler

Resource-Rich

VRE Management System

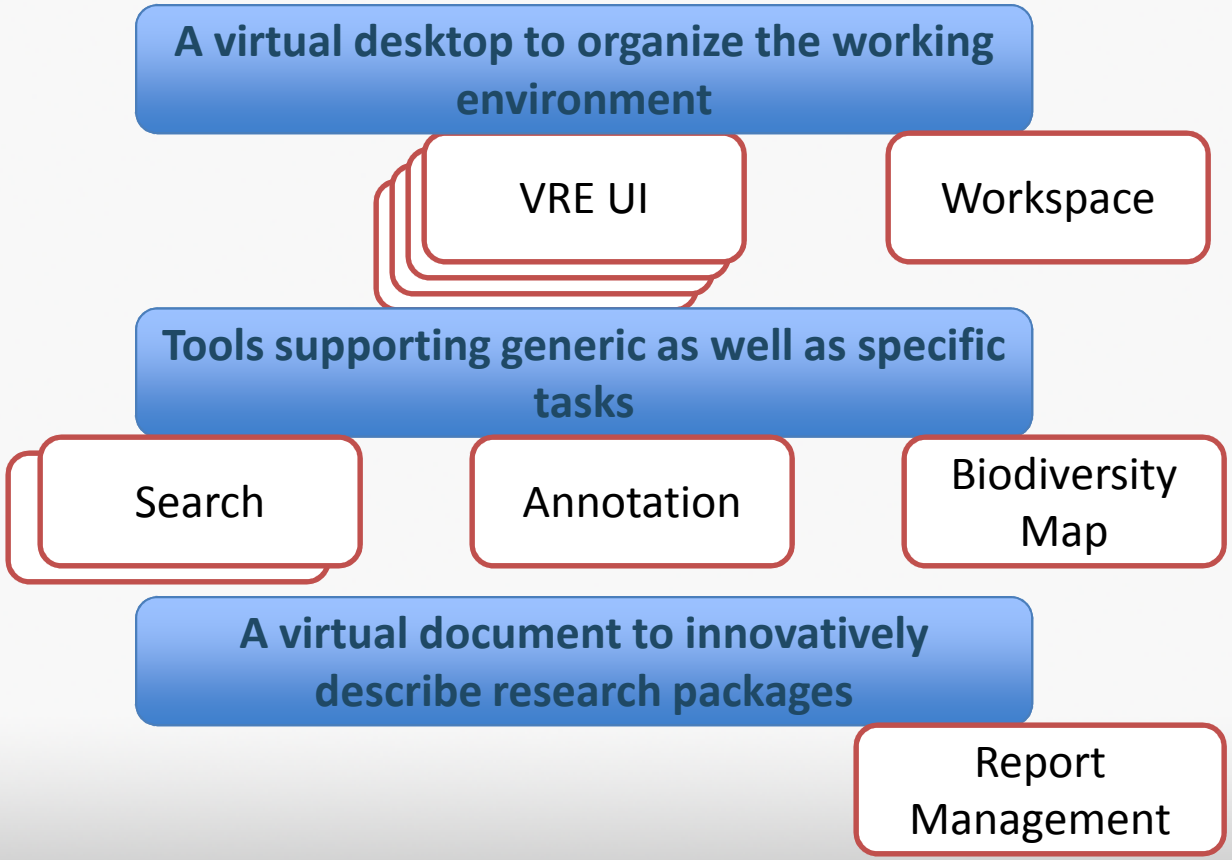
Autonomic, Extensible, and Maintainable

Building Virtual Research Environments



DL.org
Digital Library Interoperability, Best Practices and Modelling Foundations

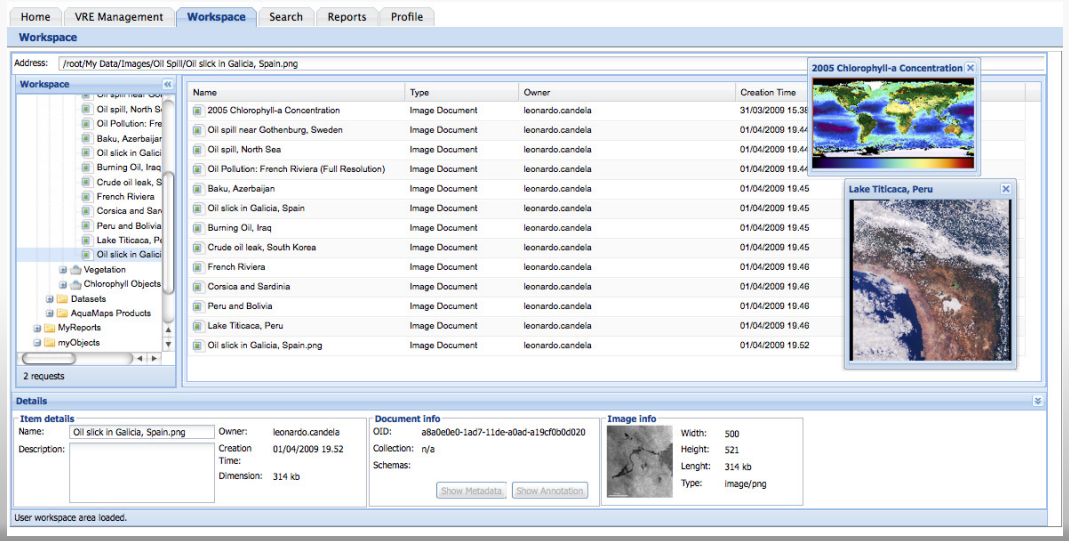
Cooperative Tools in a Nutshell



DL.org
Digital Library Interoperability, Best Practices and Modelling Foundations

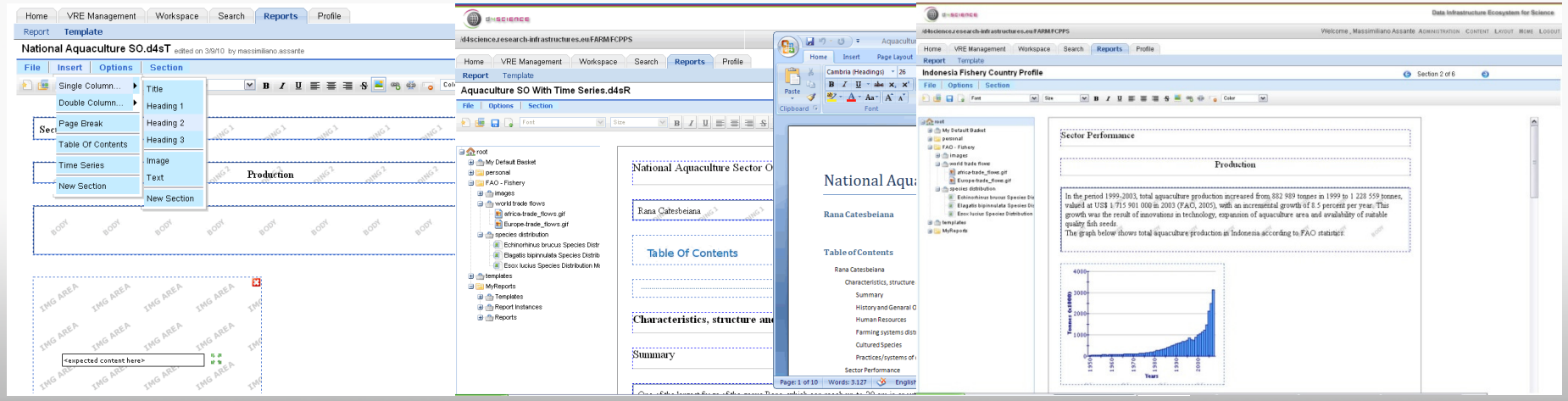
Workspace: the service

- 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



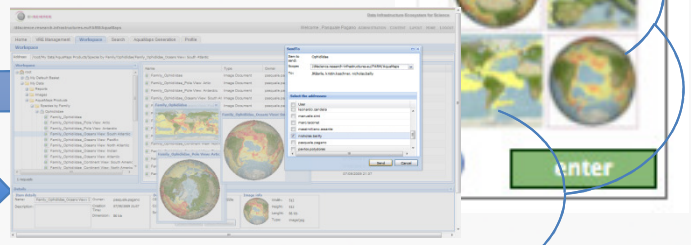
Report Management: the service

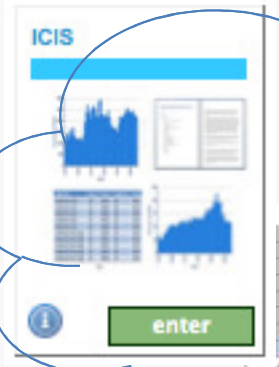
- A collaboration-oriented suite providing for
  - template-oriented, feature-rich and flexible document format definition
  - effective and infrastructure-integrated report compilation (drag & drop workspace items)
  - collaborative and distributed editing (workspace based)
  - standard-based report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

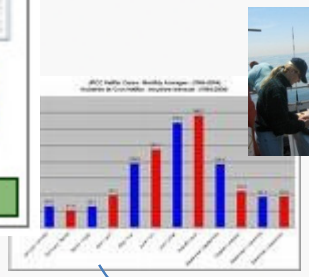


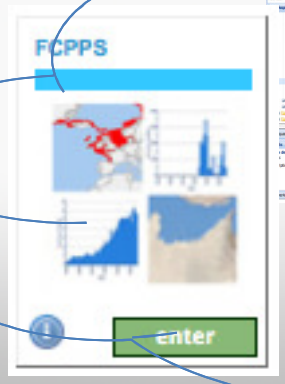
VREs, Workspaces and Report in Action

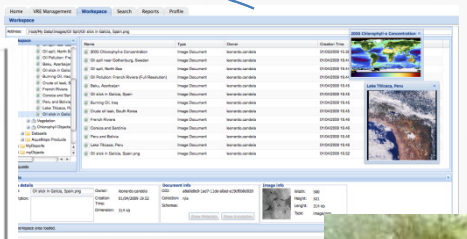
DL.org
Digital Library Interoperability, Best Practices and Modelling Foundation

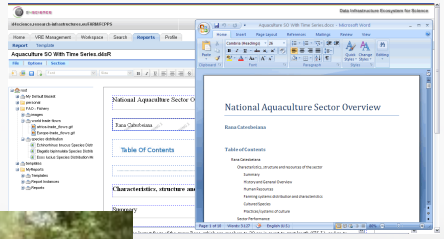












DL.org
Digital Library Interoperability, Best Practices and Modelling Foundations

gCube and Humanities: the gMan case

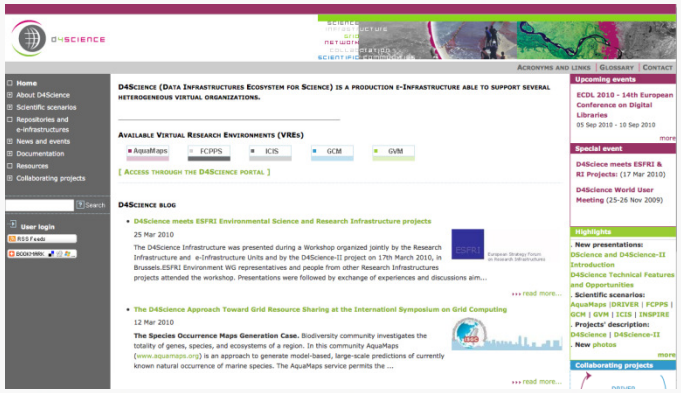
- JISC - King's College London
- Look at new ways of integrating existing data resources for Classics and add services so that research work based on integrated resources can be published
- Data sources
  - The Heidelberger Gesamtverzeichnis (HGV) der griechischen Papyrusurkunden Aegyptens, a collection 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"

DL.org
Digital Library Interoperability, Best Practices and Modelling Foundations

gCube / D4Science

www.gcube-system.org

www.d4science.eu



Donatella Castelli
D4Science-II Project Director
donatella.castelli@isti.cnr.it

Pasquale Pagano
D4Science-II Technical Director
pasquale.pagano@isti.cnr.it