

An Extensible Virtual Digital Libraries Generator

M. Assante **L. Candela** D. Castelli L. Frosini L. Lelii
P. Manghi A. Manzi P. Pagano M. Simi

Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" – CNR, Pisa - Italy
`name.surname@isti.cnr.it`

ECDL 2008, September 16th



Research trend

e-Science scenarios (person-centric, multidisciplinary and co-operative) face novel challenges

- highly-**evolving requirements**
- **large scale** resources and players distribution
- **heterogeneity**

... making standard development approaches often too “expensive” (and not sustainable)

- “from-scratch” development of ad-hoc solutions
- HW investment (even if intermittently needed)

The “magic” formula to reduce costs

sharing & reuse



Research trend

e-Science scenarios (person-centric, multidisciplinary and co-operative) face novel challenges

- highly-**evolving requirements**
- **large scale** resources and players distribution
- **heterogeneity**

... making standard development approaches often too “expensive” (and not sustainable)

- “from-scratch” development of ad-hoc solutions
- HW investment (even if intermittently needed)

The “magic” formula to reduce costs

sharing & reuse



Research trend

e-Science scenarios (person-centric, multidisciplinary and co-operative) face novel challenges

- highly-**evolving requirements**
- **large scale** resources and players distribution
- **heterogeneity**

... making standard development approaches often too “expensive” (and not sustainable)

- “from-scratch” development of ad-hoc solutions
- HW investment (even if intermittently needed)

The “magic” formula to reduce costs

sharing & reuse

DLMSs on [Grid-based] e-Infrastructures as enabling technologies

A physical and organisational structure based on the principle of **co-ordinate resource sharing** to serve one or more communities and support their operation

- originally focused on low-level resources, e.g. network, storage, computing
- lifting the vision into application domains, e.g. retrieval services and DLMSs

The impact is potentially non-trivial (**economy of scale**)

- bunch of (V)DLs operated on the same pool of resources
- innovative services available (bidirectional benefit)



DLMSs on [Grid-based] e-Infrastructures as enabling technologies

A physical and organisational structure based on the principle of **co-ordinate resource sharing** to serve one or more communities and support their operation

- originally focused on low-level resources, e.g. network, storage, computing
- lifting the vision into application domains, e.g. retrieval services and DLMSs

The impact is potentially non-trivial (**economy of scale**)

- bunch of (V)DLs operated on the same pool of resources
- innovative services available (bidirectional benefit)



DILIGENT(D4Science) in a nutshell

An innovative grid based e-Infrastructure serving e-Science scenarios through **Virtual Research Environments¹**, i.e. **dynamically generated** environments providing scientists with seamless access to all the need resources, regardless of their physical location

The **gCube** system

- **sharing** of (1) *computational resources*, (2) *structured data*, and (3) *application services*
- **service-orientation**, 3 logical tiers (Core + Info Mgmt + Presentation)
- 60 web services, 44 software libraries, 33 portlets

¹a.k.a. **Virtual Digital Libraries** or Collaboratories

DILIGENT(D4Science) in a nutshell

An innovative grid based e-Infrastructure serving e-Science scenarios through **Virtual Research Environments**¹, i.e. **dynamically generated** environments providing scientists with seamless access to all the need resources, regardless of their physical location

The **gCube** system

- **sharing** of (1) *computational resources*, (2) *structured data*, and (3) *application services*
- **service-orientation**, 3 logical tiers (Core + Info Mgmt + Presentation)
- 60 web services, 44 software libraries, 33 portlets

¹a.k.a. **Virtual Digital Libraries** or Collaboratories

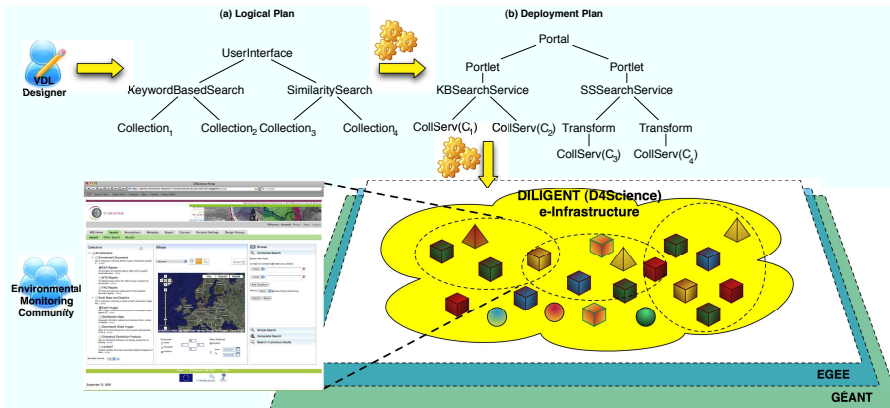


Goal

Equip the gCube system with a **VDL Generator Framework**, i.e. a service supporting VDLs definition and operation on e-Infrastructure

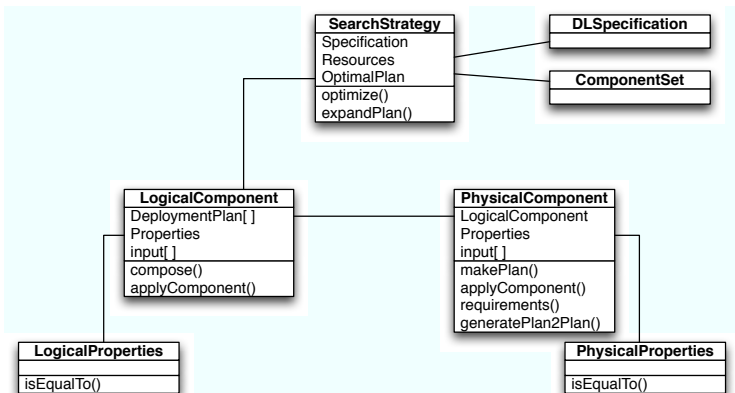
- **open** and **modular**
 - cope with diverse, autonomous, and pluggable elements/resources
- suited to **maximise infrastructure exploitation and support**
 - avoid complexities and errors while defining VDL
 - hide complexities while deploying VDL
 - promote optimal elements (and resources) consumption (avoid mis-utilisation and misuse)
 - generation strategy solutions on-board
 - operational framework for experimentation with novel technologies and solutions

VDLs Generation



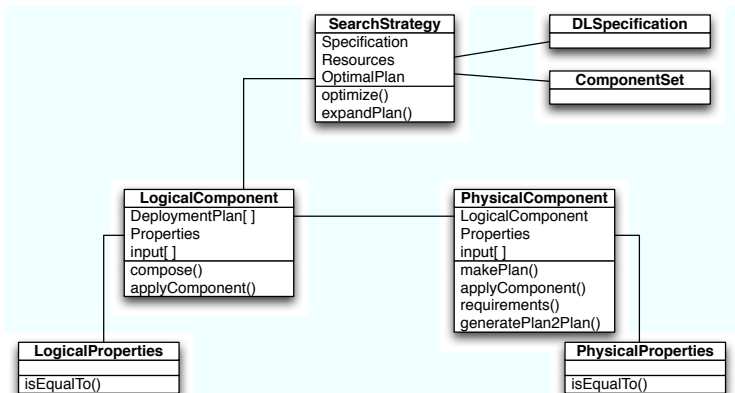
Overall View

Logical Plans: trees of logical components (abstract class implementations)



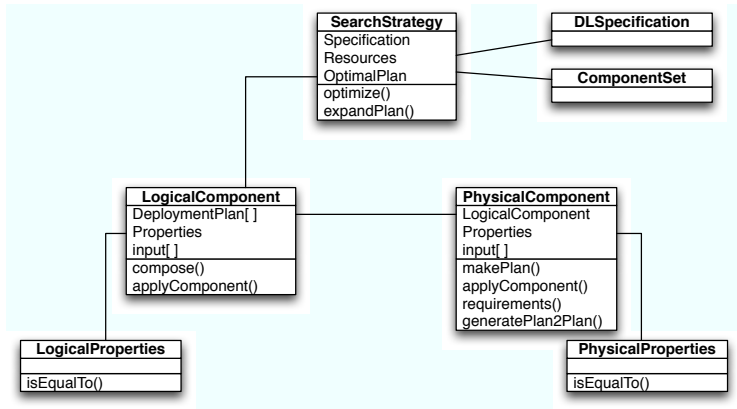
Overall View

Deployment Plans: threes of physical components (abstract class implementations)



Overall View

Search Strategy: strategy to produce deployment plans



Representing and Generating Logical Plans

- one `LogicalComponent` derived class for each logical part of a DL
 - e.g. a collection, a functionality, a user interface part
- two key methods
 - `compose` – to produce a new logical plan (inherited)
 - `applyComponent` – to decide about the component applicability (per component)
- one `LogicalProperties` to capture the peculiarities of the current plan



Representing and Generating Deployment Plans

- one `PhysicalComponent` derived class for each architectural part of a DL system
 - e.g. a service, a portlet
- two key methods
 - `compose` – to produce a new deployment plan (inherited)
 - `applyComponent` – to decide about the component applicability (per component)
- one `PhysicalProperties` to capture the peculiarities of the current deployment plan
- *enforcers*
 - e.g. data transformation, supporting service
 - `requirements` – to declare per component applicability needs
 - `generatePlan2Plan` – to enrich the input plan with one of the available *enforcers*

The Search Strategy

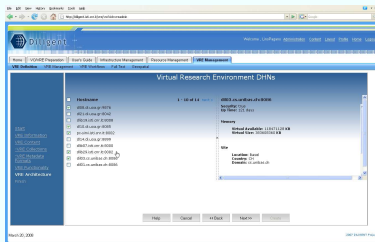
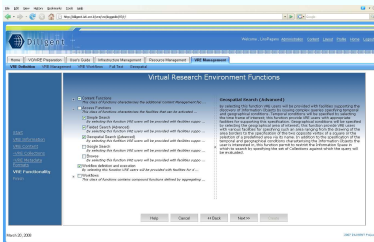
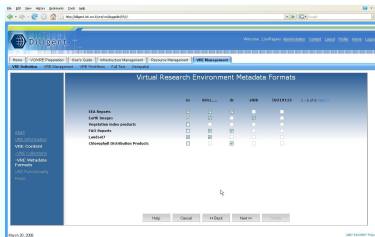
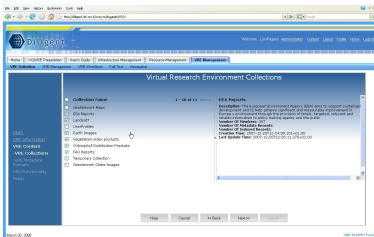
- entirely implemented in terms of *abstract* classes and methods
 - i.e. search strategy is domain-agnostic
- two key methods (dynamic programming)
 - `optimize` – to orchestrate the whole strategy
 - `expandPlan` – to perform a single step of optimisation strategy
- per step property-based (e.g. cost) pruning to drop sub-optimal plans



gCube Components

- examples of Logical Components
 - `Collection` to represent each data source
 - `Search` to represent the relative functionality applied to an information space
- examples of Physical Components
 - `CollectionService` to represent the data source service
 - `GeoSearch` to the geo-referenced search service
- examples of enforcers
 - `QueryPlanner` and `GeoIndexLookup` to support the `GeoSearch`
 - `gHN` for **dynamic service deployment**
 - `factory service` for **dynamic resource deployment**

The Wizard



Summary

e-Science scenarios demand for **infrastructure-oriented approaches** to guarantee low-costs and sustainability

- the higher initial development cost than traditional ad-hoc solutions is well **repaid by the long-term scale of adoption and maintenance**

The **gCube VDL Generator Framework**

- is **open** thus to guarantee the easy of use/adaptation in unexpected scenarios
- makes user-friendly (wizard-based) a difficult task

<http://www.d4science.eu>

<http://www.gcube-system.org>

Demo in the afternoon

Summary

e-Science scenarios demand for **infrastructure-oriented approaches** to guarantee low-costs and sustainability

- the higher initial development cost than traditional ad-hoc solutions is well **repaid by the long-term scale of adoption and maintenance**

The **gCube VDL Generator Framework**

- is **open** thus to guarantee the easy of use/adaptation in unexpected scenarios
- makes user-friendly (wizard-based) a difficult task

<http://www.d4science.eu>

<http://www.gcube-system.org>

Demo in the afternoon