

On-Demand Service Deployment and Process Support in e-Science DLs: the DILIGENT Experience

Leonardo Candela¹ Donatella Castelli¹ Christoph Langguth²
Pasquale Pagano¹ Heiko Schuldt² Manuele Simi¹
Laura Voicu²

¹Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" - CNR - Pisa, Italy
{candela|castelli|pagano|simi}@isti.cnr.it

²University of Basel – Department of Computer Science - Basel, Switzerland
{christoph.langguth|heiko.schuldt|laura.voicu}@unibas.ch



Diligent

e-Science Community and DLs

Nowadays research is a **multidisciplinary** and **co-operative** effort

- virtual research organisations not necessarily
 - aggregated for long periods
 - having great economic power and DL expertise
- DLs as supporting tools
 - seamless access to their **resources**, i.e. content and services
- from DLs to Knowledge Commons (DELOS view)
 - from content-centric to person-centric
 - from information storage and retrieval to communication and collaboration supporting tool
 - from centrally-located text to distributed and heterogeneous data sources



e-Science Community and DLs

Nowadays research is a **multidisciplinary** and **co-operative** effort

- virtual research organisations not necessarily
 - aggregated for long periods
 - having great economic power and DL expertise
- DLs as supporting tools
 - seamless access to their **resources**, i.e. content and services
- from DLs to Knowledge Commons (DELOS view)
 - from content-centric to person-centric
 - from information storage and retrieval to communication and collaboration supporting tool
 - from centrally-located text to distributed and heterogeneous data sources



e-Science Community and DLs

Nowadays research is a **multidisciplinary** and **co-operative** effort

- virtual research organisations not necessarily
 - aggregated for long periods
 - having great economic power and DL expertise
- DLs as supporting tools
 - seamless access to their **resources**, i.e. content and services
- from DLs to Knowledge Commons (DELOS view)
 - from content-centric to person-centric
 - from information storage and retrieval to communication and collaboration supporting tool
 - from centrally-located text to distributed and heterogeneous data sources



e-Science Community and DLs

Nowadays research is a **multidisciplinary** and **co-operative** effort

- virtual research organisations not necessarily
 - aggregated for long periods
 - having great economic power and DL expertise
- DLs as supporting tools
 - seamless access to their **resources**, i.e. content and services
- from DLs to Knowledge Commons (DELOS view)
 - from content-centric to person-centric
 - from information storage and retrieval to communication and collaboration supporting tool
 - from centrally-located text to distributed and heterogeneous data sources



e-Science Community: a Scenario

A naval accident provokes a large oil spill dump. The local environment agency

- sets up a team of experts dedicated to manage the disaster
- sets up an integrated environment providing seamless access to
 - heterogeneous information sources (satellite images, reports, sensor data)
 - needed functionality (e.g. search related events, estimate oil volumes, simulate action results)
 - computing resources to operate it

i.e. a DL for supporting the “crisis” in an efficient and effective way



e-Science Community: a Scenario

A naval accident provokes a large oil spill dump. The local environment agency

- sets up a team of experts dedicated to manage the disaster
- sets up an integrated environment providing seamless access to
 - heterogeneous information sources (satellite images, reports, sensor data)
 - needed functionality (e.g. search related events, estimate oil volumes, simulate action results)
 - computing resources to operate it

i.e. a DL for supporting the “crisis” in an efficient and effective way



e-Science Community: a Scenario

A naval accident provokes a large oil spill dump. The local environment agency

- sets up a team of experts dedicated to manage the disaster
- sets up an integrated environment providing seamless access to
 - heterogeneous information sources (satellite images, reports, sensor data)
 - needed functionality (e.g. search related events, estimate oil volumes, simulate action results)
 - computing resources to operate it

i.e. a DL for supporting the “crisis” in an efficient and effective way



e-Science Community: a Scenario

A naval accident provokes a large oil spill dump. The local environment agency

- sets up a team of experts dedicated to manage the disaster
- sets up an integrated environment providing seamless access to
 - heterogeneous information sources (satellite images, reports, sensor data)
 - needed functionality (e.g. search related events, estimate oil volumes, simulate action results)
 - computing resources to operate it

i.e. a DL for supporting the “crisis” in an efficient and effective way



DLs for e-Science Community

Current DL development model . . .

- “from scratch” development
- ad hoc resources provisioning

. . . is not appropriate! [takes time and money]

New DL development model

- DL built by dynamically aggregating the needed resources
- new functionality combined in user defined workflows

i.e., DILIGENT



Diligent

DLs for e-Science Community

Current DL development model . . .

- “from scratch” development
- ad hoc resources provisioning

. . . is not appropriate! [takes time and money]

New DL development model

- DL built by dynamically aggregating the needed resources
- new functionality combined in user defined workflows

i.e., DILIGENT



Diligent

DLs for e-Science Community

Current DL development model . . .

- “from scratch” development
- ad hoc resources provisioning

. . . is not appropriate! [takes time and money]

New DL development model

- DL built by dynamically aggregating the needed resources
- new functionality combined in user defined workflows

i.e., DILIGENT



Diligent

DILIGENT: An Infrastructure Supporting e-Science DLs

Provide an advanced test-bed DL infrastructure allowing e-Scientists to access shared knowledge and to collaborate in a secure, co-ordinated, dynamic and cost-effective way

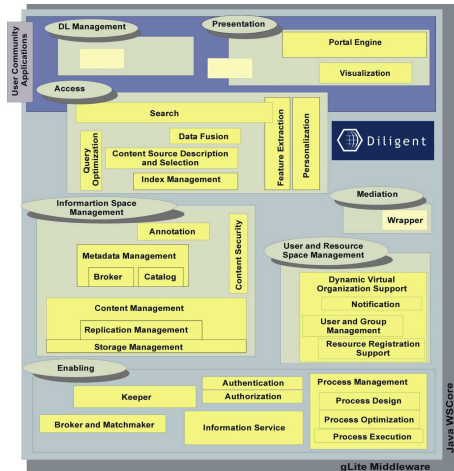
- combines **Grid** and **Digital Library** technologies
- builds on top of Enabling Grid for E-science (EGEE) project
 - aims to deliver the largest worldwide service Grid Infrastructure (more than 90 partners, 32 countries)
- acts as *DL Broker* over a pool of **shared resources**
 - *collections*, i.e. set of information objects
 - *services*, i.e. software entities implementing a functionality
 - *hosting nodes*, i.e. networked entities capable to host collections and services
 - *gLite resources*, i.e. computing and storage facilities

promoting their usage **on-demand**



Service Oriented Architecture over Grid Framework

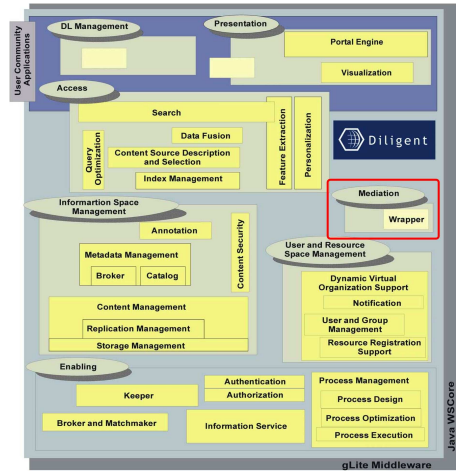
- DL as a federation of services accessed through a portal
- New functionality delivered via workflows of services
- Application framework
 - gLite Grid middleware
 - WSRF specification (Globus toolkit implementation)



Services Overview

Mediation

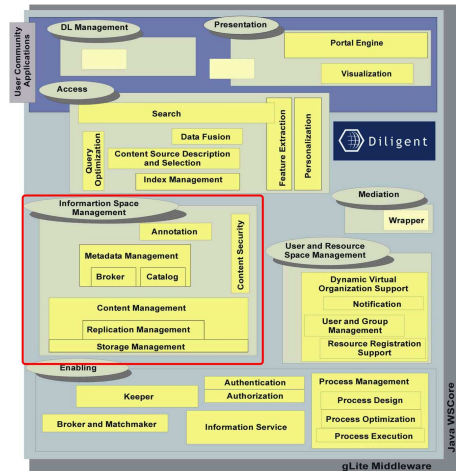
- making external objects available as DILIGENT information objects via wrappers



Services Overview

Information Space Management

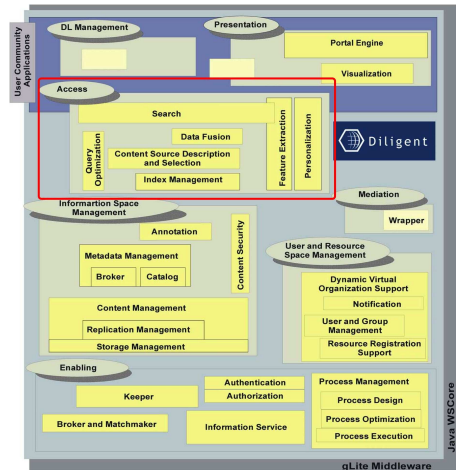
- composite objects
 - local and gLite storage facilities
 - replica management
- metadata in multiple format
- annotation
- secure content
 - watermarking and encryption



Services Overview

Access

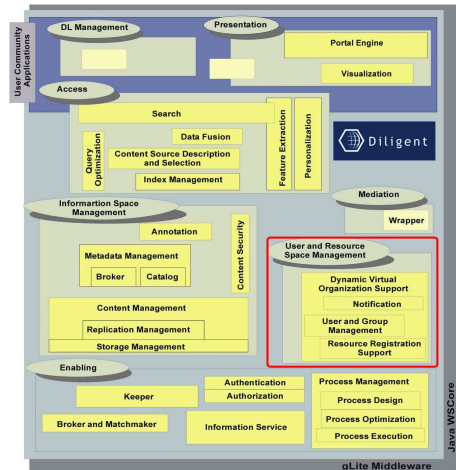
- powerful search
 - various types of indexes
 - multimedia and multi-type feature extraction
 - automatic source description and selection
 - data fusion
 - personalisation
- ... co-ordinated via workflow support



Services Overview

User & Resource Management

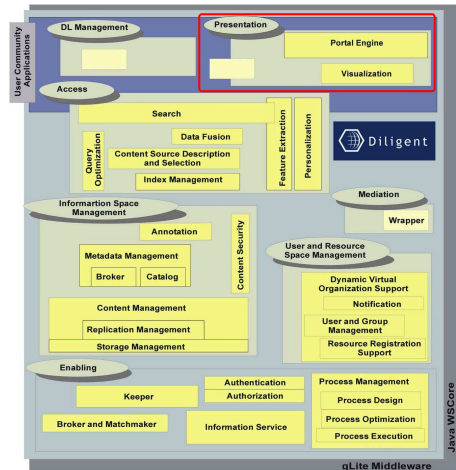
- user and resource registration and validation
- notification support
- VO mechanism for ensuring DL resource sharing
- AuthN & AuthZ (X.509 and GSI)



Services Overview

Presentation

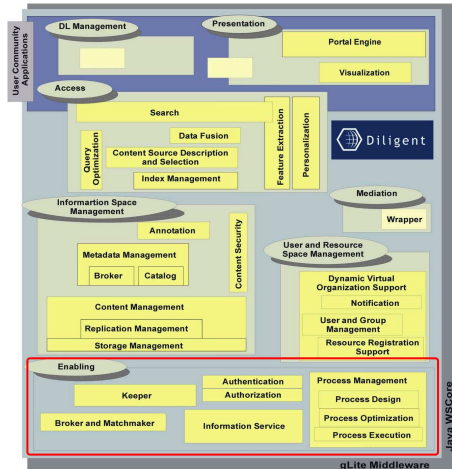
- user oriented access point to the DL (portal)
- plug & play of community specific tools
- GridSphere + standards (JSR168 and WSRP)



Services Overview

Enabling

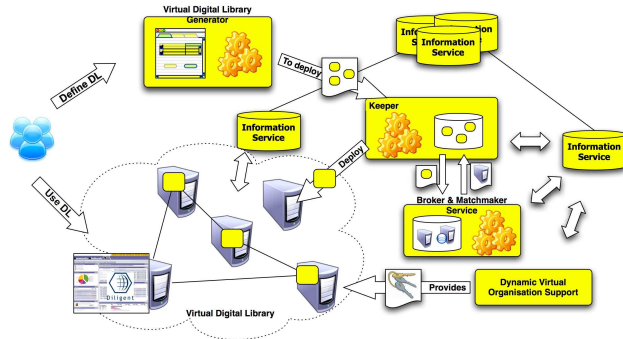
- monitoring and discovering DILIGENT resources
- optimal usage of available resources
- DL federation creation and maintenance (dynamic deployment)
- workflows design and verification
- workflows execution



DLs Creation & Maintenance: the Keeper Service

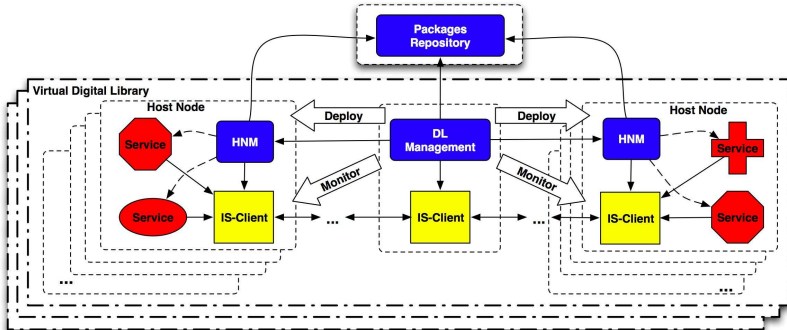
Manages DILIGENT DLs

- deploys new services instances
- monitors DL status
- guarantees DL operation
- maintains the requires QoS



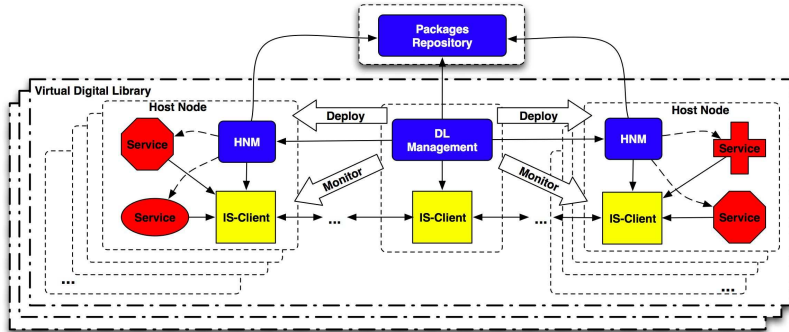
Keeper Service: the Architecture

Distributed architecture



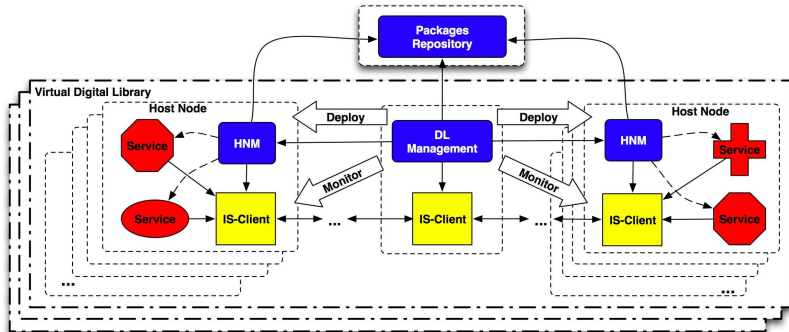
Keeper Service: the Architecture

Centralised storage of DILIGENT software



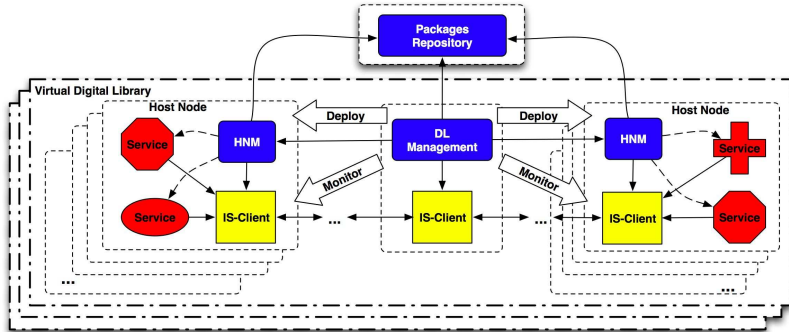
Keeper Service: the Architecture

Centralised per DL orchestration



Keeper Service: the Architecture

Distributed hosting node management



Information Service

Supports resources **discovery** and **monitoring**

- gathers, stores and supply information about the resources constituting DILIGENT and needed to the other services
 - no pre-established paths
- XML-based Resource Profiles
- Push and pull modalities, i.e. query and subscribe/notify both in producing and consuming information

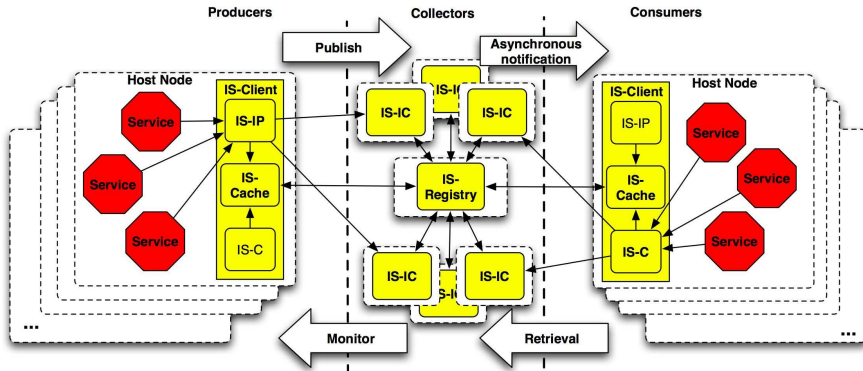
Issues

- large numbers (many resources, many users)
- dynamic information
- communication overhead
- VO/DL focused views



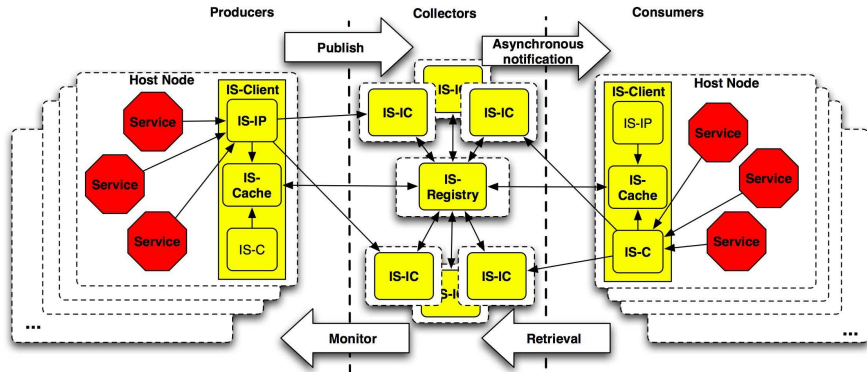
Information Service: the Architecture

Grid Monitoring Architecture (GMA) Approach



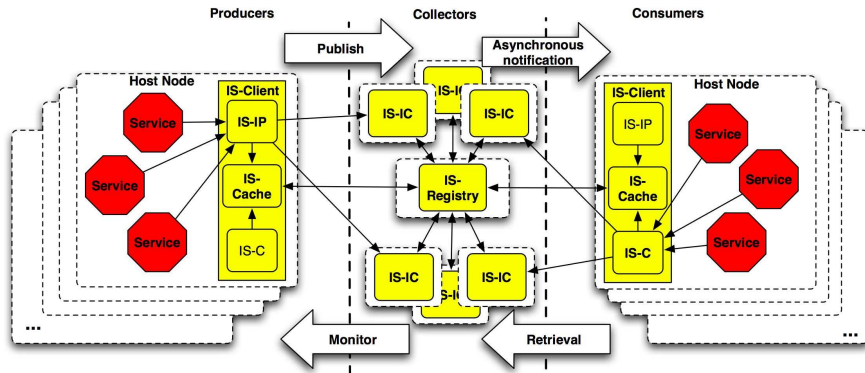
Information Service: the Architecture

Lightweight Client Library



Information Service: the Architecture

Replicated & Distributed Collectors



Process Design and Validation

Implement complex services by combining existing ones, a.k.a. **programming in the large**

- control flow, i.e. services invocation order
- data flow, i.e. data items transfer between services
- transactional behaviour and execution guarantees for concurrency and failure handling

XML, SOAP, and WSDL as technologies for service description and invocation

BPEL as foundation for process specification



Process Design and Validation

Implement complex services by combining existing ones, a.k.a.
programming in the large

- control flow, i.e. services invocation order
- data flow, i.e. data items transfer between services
- transactional behaviour and execution guarantees for concurrency and failure handling

XML, SOAP, and WSDL as technologies for service description and invocation

BPEL as foundation for process specification



Process Design and Validation

Implement complex services by combining existing ones, a.k.a. **programming in the large**

- control flow, i.e. services invocation order
- data flow, i.e. data items transfer between services
- transactional behaviour and execution guarantees for concurrency and failure handling

XML, SOAP, and WSDL as technologies for service description and invocation

BPEL as foundation for process specification



Process Design and Validation: the Tool

The screenshot displays the 'Applet' tool interface. The main workspace shows a process flow diagram with the following steps: 'initialize' → 'login' → 'get session token' → 'submit' → 'get task taken' → 'fetch result name' → 'query status' → 'get filename' → 'get ftp endpoint' → 'assign ftp var' → 'store into CMG'. A 'step' activity is also shown connected to 'query status'.

The right-hand side features a 'Process Properties' panel with sections for 'Whiteboard & I/O' and 'Process Input/Output Arguments'. The 'Whiteboard & I/O' section includes a table for defining whiteboard entries:

Name	Type
ProcessorEPR	String
SessionEPR	String
ResponseTaskIdEPR	String

The bottom section contains 'Activity Properties' and 'Invocation Properties & I/O'. The 'Invocation Properties & I/O' section includes a table for defining activity input arguments:

Tag	Value
SOAP_Workaround	true
SOAP_Header_Count	1

The 'Activity Properties' section includes fields for 'Name' (query status) and 'Fill from WSDL'.



Process Execution

Process specification combines “types” of services, execution maps them in **concrete service calls**

- service instances discovered through the Information Service

To profit from resource availability and promote their optimal use a truly distributed execution infrastructure is needed

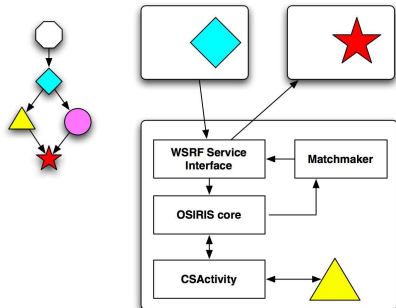
- process is executed over an overlay network of CSEngines
- routing strategy locally decided



Process Execution: Concrete Components

Build on top of OSIRIS

- provide SOAP interface to CSEngine
- orchestrates execution by identifying the CSAActivity to be done
- executes the workflow task, BPEL `<invoke>` and `<assign>`
- identify the “best” next node



The ImpECt Scenario

ImpECt goal is to support Earth Observation community in managing the reports production task

- have seamless access to heterogeneous information sources
 - environmental conventions and news
 - satellite images
 - raw data gathered from satellite sensors
- dynamically create the DL environment for serving their task
 - aggregating the relevant information sources
 - aggregating the needed services, computing, and storage resources
 - appropriately combining services to deliver new and customised functionality



The ImpEct Scenario: Living Reports

**International Report on
 Mediterranean Sea Chlorophyll Distribution during year 2003**

1. Scientific and Societal Concerns
 Any scheme to monitor the ocean biota and their environment must strive to address the major scientific and societal concerns of the day pertaining to marine life. This section summarizes some major concerns that emerged during discussions at the meeting. Many other concerns could have been included, but space precludes a complete listing of concerns.

1.1. Biodiversity and Conservation
 Marine biodiversity is not easy to assess and is generally poorly known. There are many complicating factors, including a three-dimensional, fluid, mobile environment, its vastness, and its challenging depths. Away from shore, primary producers and primary grazers are usually small, drifting forms that undergo spatial variability and seasonal changes. The larger invertebrate grazers have a range of life history stages, often with planktonic and benthic phases. Many large animals are migratory. Ocean habitats can be linked by the dispersal of planktonic larvae, and in this way, the systems can be interconnected over a distance.

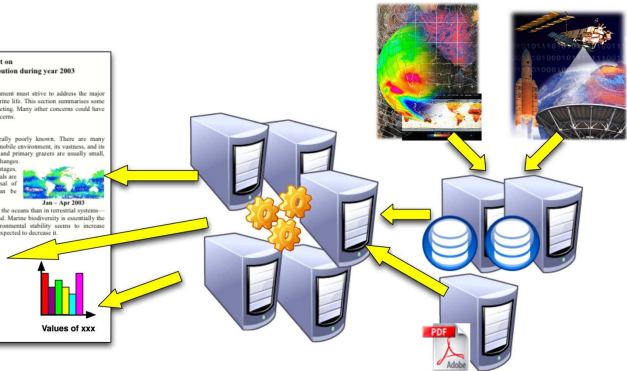
Jan - Apr 2003

Finally, the higher-order diversity of life is much greater in the oceans than in terrestrial systems—there are 13 unique phyla in the oceans and only one on land. Marine biodiversity is essentially the evolutionary history of life. In general, long-term environmental stability seems to increase biodiversity and, conversely, global climate change can be expected to decrease it.

	X1	X2	X3	X4	X5	X6	X7	X8	X9
Y1	12	13	15	26	11	34	45	45	54
Y2	32	12	46	67	21	22	44	12	44
Y3	23	33	56	77	32	44	12	55	33
Y4	44	34	12	55	34	45	12	22	44

Measures of yyy

Values of xxx



Summary

e-Science DLs have a different nature than traditional DLs

- they must rapidly adapt to the evolving requirements of their communities

A new development model

- supporting dynamic DL creation and management by resources aggregation and dynamic deployment
- having controlled sharing as a foundation
- easing the creation of complex and personalised functionality

is needed

DILIGENT implements this model

<http://www.diligentproject.org>



Diligent

Summary

e-Science DLs have a different nature than traditional DLs

- they must rapidly adapt to the evolving requirements of their communities

A new development model

- supporting dynamic DL creation and management by resources aggregation and dynamic deployment
- having controlled sharing as a foundation
- easing the creation of complex and personalised functionality

is needed

DILIGENT implements this model

<http://www.diligentproject.org>



Diligent

Summary

e-Science DLs have a different nature than traditional DLs

- they must rapidly adapt to the evolving requirements of their communities

A new development model

- supporting dynamic DL creation and management by resources aggregation and dynamic deployment
- having controlled sharing as a foundation
- easing the creation of complex and personalised functionality

is needed

DILIGENT implements this model

<http://www.diligentproject.org>



Diligent

Summary

e-Science DLs have a different nature than traditional DLs

- they must rapidly adapt to the evolving requirements of their communities

A new development model

- supporting dynamic DL creation and management by resources aggregation and dynamic deployment
- having controlled sharing as a foundation
- easing the creation of complex and personalised functionality

is needed

DILIGENT implements this model

<http://www.diligentproject.org>



Diligent