

Digital Libraries: a worldwide distributed application

Donatella Castelli
ISTI-CNR
Pisa (Italy)

 Large widely “distributed” DLs

 DILIGENT: A Digital Library Infrastructure on Grid ENabled Technology

The new Alexandria Library



First EUMEDGRID Project Conference, 15th September 2006, Rome

Large distributed DLs: NSDL

National Science Digital Library (NSDL)

<http://nsdl.org/index.php>

- Funding Agency: National Science Foundation
- Start date: 2000



- Mission:

To provide organized access to high quality resources and tools that support innovations in teaching and learning at all levels of science, technology, engineering, and mathematics (STEM) education.

Large distributed DLs: NSDL [cont]

- **Content providers**

Other digital libraries, NSF-funded projects, and NSDL-reviewed web sites

- **Consumers**

Primarily educators, but anyone can access NSDL.org and search the library at no cost

Large distributed DLs: TEL

The European Library

http://libraries.theeuropeanlibrary.org/aboutus_en.htm



- Funding Agency: EU - CENL
- Start date: 2001
- Mission: *The European Library exists to open up the universe of knowledge, information and cultures of all Europe's national libraries.*
- Content providers:
45 national libraries of Europe - books, magazines, journals.... - both digital and non-digital

Large distributed DLs: DRIVER

Digital Repository Infrastructure Vision for European Research

(<http://www.driver-repository.eu/>=)



- **Funding Agency:** EU Research Networking Test-beds
- **Start date:** June 2006
- **Duration:** 18 months
- **Objective:**
To develop a test-bed for integrating existing national, regional and/or thematic repositories in order to create a production-quality European infrastructure

Large distributed DLs: DRIVER

- **Initial set of content providers: ~ 50 institutional repositories**
 - ◆ UK(SHERPA), The Netherland (DARE) + Germany(BASE)
Italy, France, Greece, Belgium, Poland (repositories managed by partner organisations)
 - ◆ Mainly texts, a small portion of multimedia objects

i2010 Digital Libraries Initiative

“The European Digital Library is a flagship project of the Commission's overall strategy to boost the digital economy, the i2010 initiative”

Brussels, 25 August 2006

<http://europa.eu.int/rapid/pressReleasesAction.do?reference=IP/06/1124&format=HTML&aged=0&language=EN&guiLanguage=fr>

“The European Commission today urged EU Member States to set up large-scale digitisation facilities, so as to accelerate the process of getting Europe's cultural heritage on line via the European digital library”

“By 2008, two million books, films, photographs, manuscripts, and other cultural works will be accessible through the European Digital Library. This figure will grow to at least six million by 2010, but is expected to be much higher as, by then, potentially every library, archive and museum in Europe will be able to link its digital content to the European Digital Library.”

Knowledge Infrastructures for Scientific Information

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, October 2003

- Institutions should implement a policy to:
 1. require their researchers to deposit a copy of all their published articles in an open access repository
 2. encourage their researchers to publish their research articles in open access journals where a suitable journal exists (and provide the support to enable that to happen)
- Up to now 140 organizations have signed it

European Infrastructure for Repositories of Scientific Information:

<http://cordis.europa.eu/ist/rn/ri-cnd/wshop-080606.htm>

“A knowledge layer building on the established network and grid layers”

First EUMEDGRID Project Conference, 15th September 2006, Rome

Two important observations

- Digital information objects are not necessarily “digital analogous” of physical documents
- DLs are expensive applications



Diligent

A **D**igital **L**ibrary Infrastructure
on **G**rid **E**Nabled **T**echnology



Information Society
Technologies

General Information

A Digital Library Infrastructure on Grid Enabled Technology

(<http://www.diligentproject.org>)

- **Start date:** September 2004
- **Duration:** 3 years
- **Effort:** 1100 p/m
- **Cost:** 9.8 M Euro
- **EU funding:** 6.3 M Euro

Objective

Create a **Test-bed Digital Library Infrastructure** that allows members of dynamic virtual organizations to create on-demand transient virtual digital libraries based on shared computational, storage, multimedia, multi-type content and application resources

Consumers



DILIGENT DL infrastructure

Service A

Service B

Service C

DLCreation
service

Service D

Service E

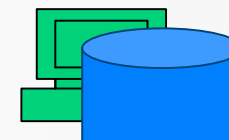
Producers



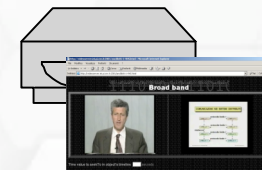
3D processing



simulation



Feature
extraction

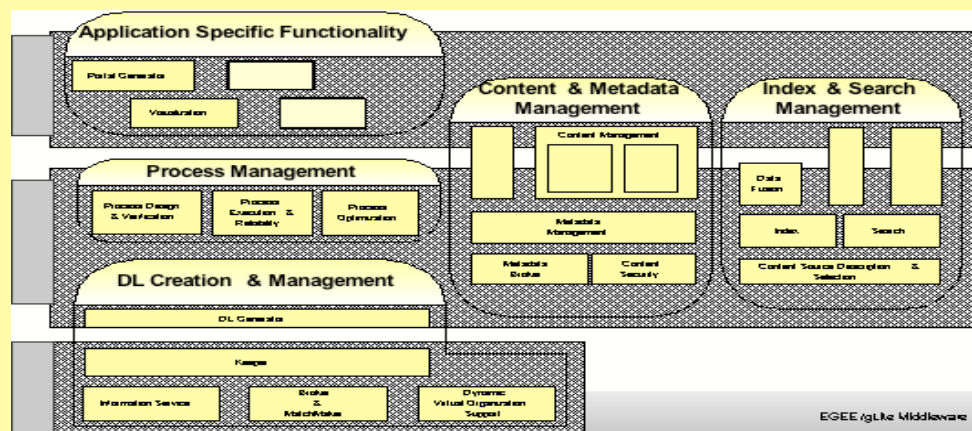


Speech
recognition

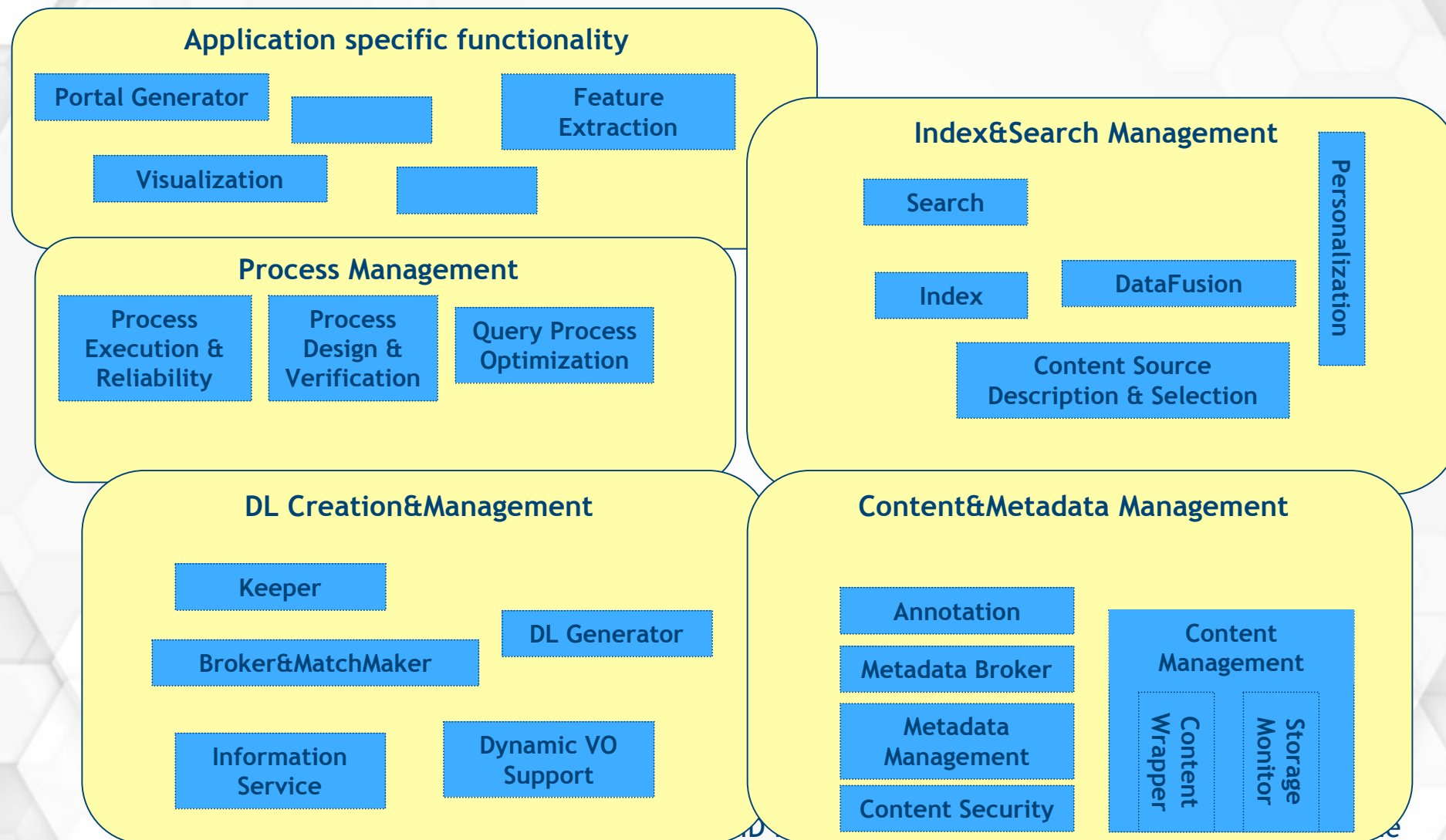
Grid main exploited features

- Controlled sharing of resources
- High computing and storage capabilities for the handling of a wide variety of information objects

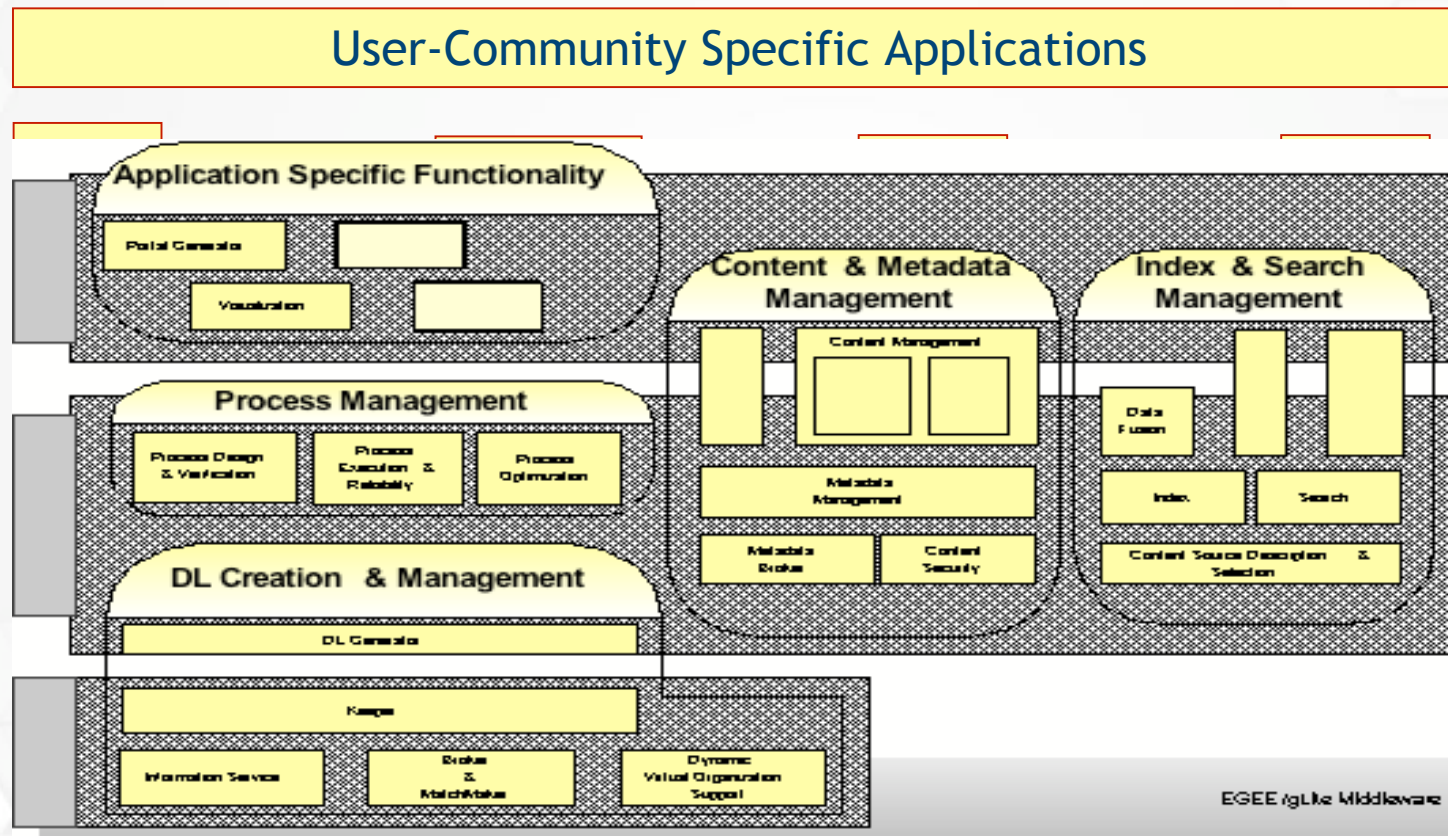
DILIGENT Architecture



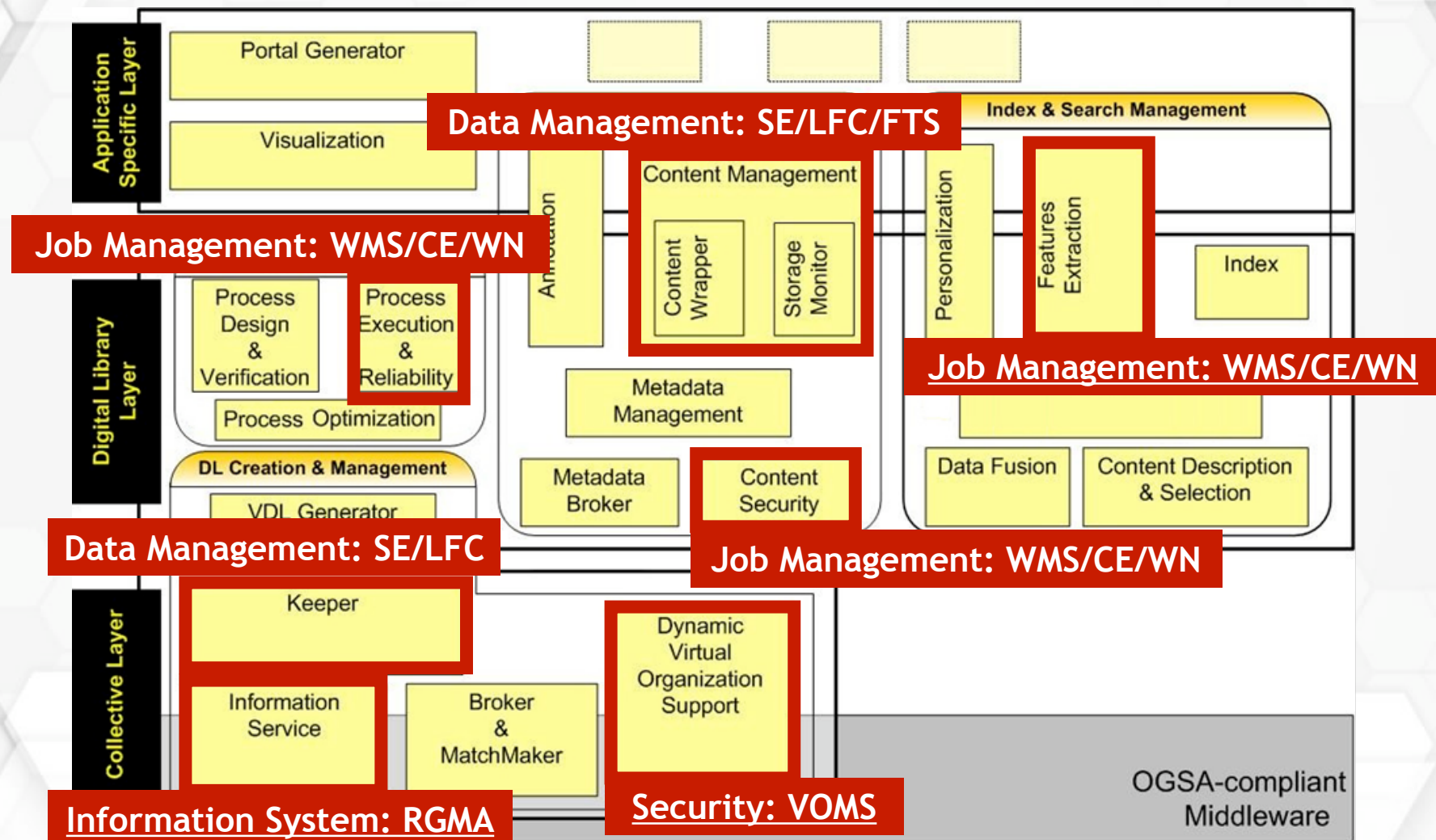
Diligent functionality decomposition



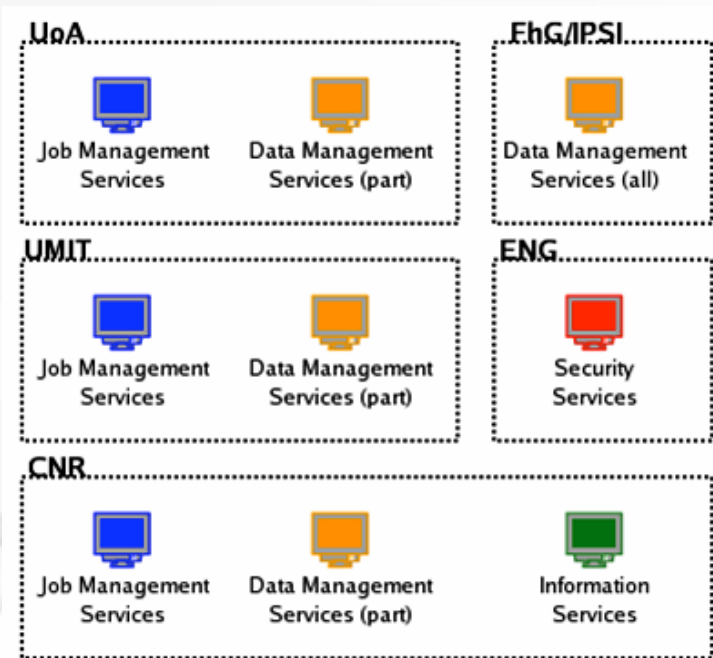
Grid framework



gLite in DILIGENT



- 2 independent infrastructures up and running all gLite services:
 - Development infrastructure
 - Testing infrastructure
- 6 sites in Athens, Budapest, Darmstadt, Pisa, Basel and Rome



DILIGENT Infrastructures



- EGEE pre-production infrastructure: two sites maintained by ESA-ESRIN and CNR-ISTI (34 CPUs and 3,35 TB)-**DILIGENT VO**

The ImpECt Scenario

Implementation of Environmental Conventions

- User community represented by the European Space Agency
- Characteristics:
 - ◆ well-established tradition in exploitation of new technologies
 - ◆ wide variety of content types (maps, satellite images, etc.)
 - ◆ very large, dynamic data sets
- Earth Science Domain DLs
 - ◆ preparation of periodical reports
 - ◆ creation of decision supporting mechanisms in case of environmental accidents

The ARTE Scenario

- User community represented by:
 - ◆ Scuola Normale Superiore
 - ◆ RAI Radiotelevisione Italiana
- Characteristics
 - ◆ IT technology exploitation still in infancy
 - ◆ multidisciplinary collaborative research
 - ◆ image based retrieval/semantic analysis of images
- Teaching and e-Learning in Humanities
 - ◆ organization of courses
 - ◆ support multidisciplinary research revolving around images

Conclusions

Not all the DL applications need to rely on a Grid infrastructure

however,

the more we experiment, the more interesting, not initially foreseen, DL applications emerge (e.g. ARCHEOGRID, FAO)

The availability of a large Grid production infrastructure is vital to foster these new DL applications and ensure their sustainability