

INDICATE

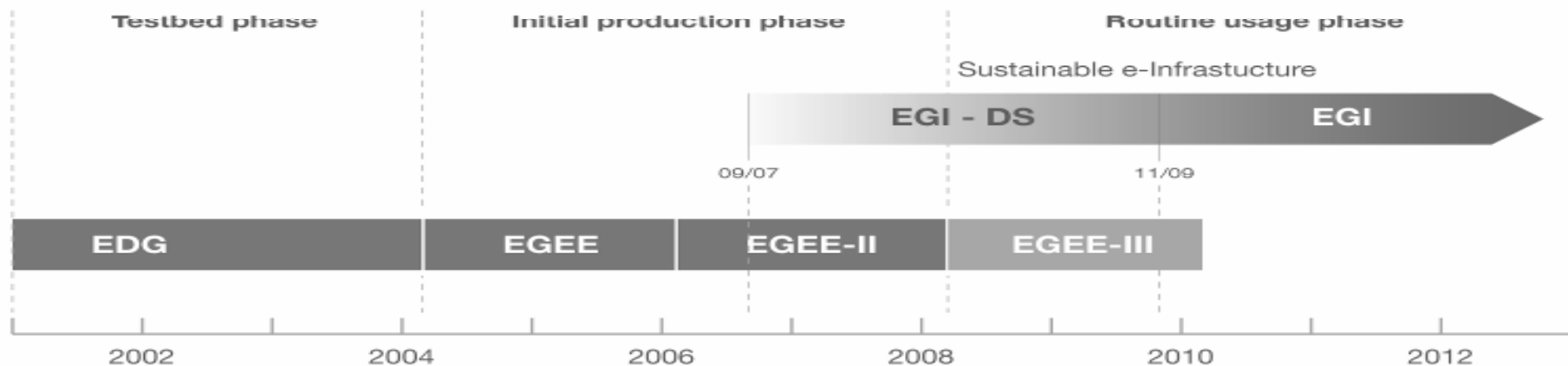
International Network for a Digital Cultural Heritage e-Infrastructure

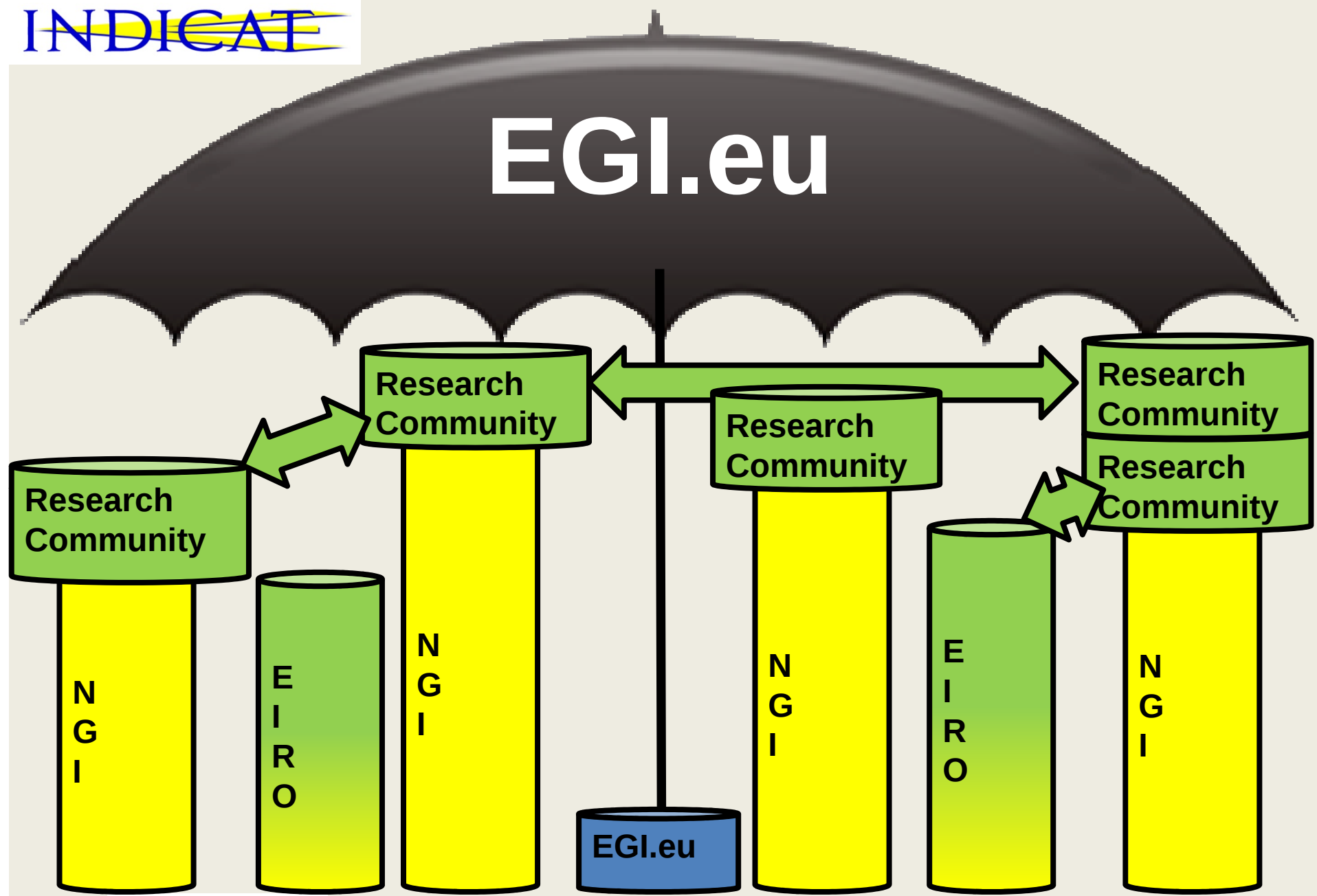
Riunione DC-NET Italia
Roma, 24 Gennaio 2011

Roberto Barbera
Consorzio COMETA

- Introduction;
- The INDICATE project:
 - Objectives, Workplan, Partners;
- COMETA and its Services for Cultural Heritage:
 - gLibrary and gLibrary Mobile;
- COMETA vs. DC-NET and INDICATE;
- The INDICATE Virtual Organisation;
- The INDICATE e-Culture Science Gateway;
- Summary and Conclusions:
 - Proposal for an INDICATE Service Challenge.

- European Data Grid (EDG) → www.edg.org
 - Explore concepts in a testbed
- Enabling Grid for E-science (EGEE) → www.eu-egee.eu
 - Moving from prototype to production
- European Grid Infrastructure (EGI) → www.egi.eu
 - Routine usage of a sustainable e-infrastructure





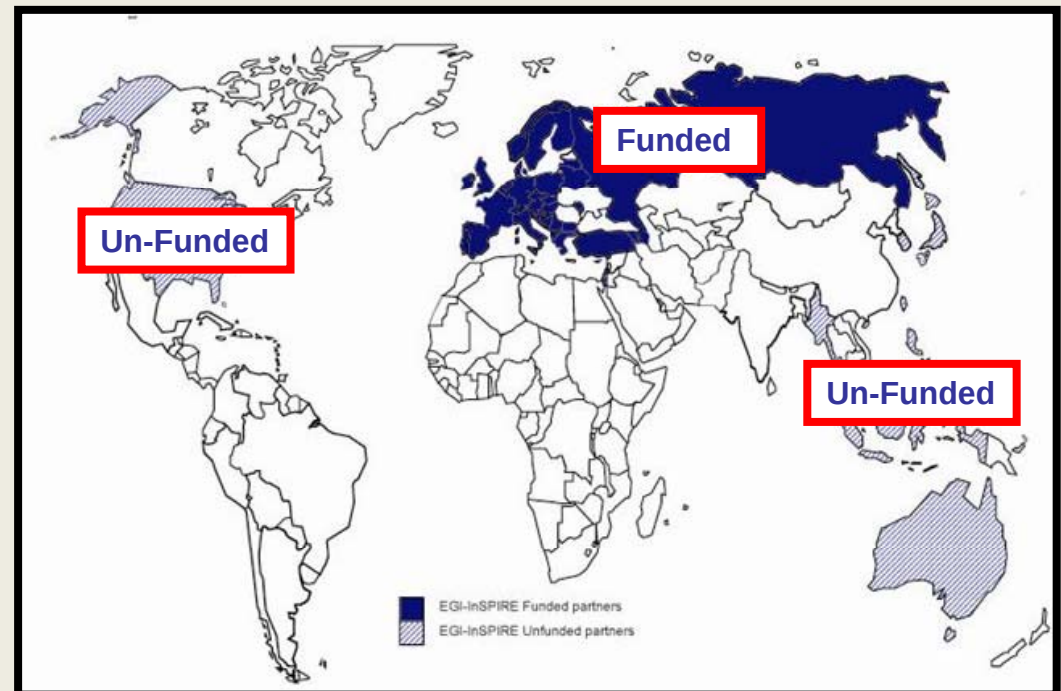
Integrated Sustainable Pan-European Infrastructure for Researchers in Europe

A 4 year project with € 25 M EC contribution

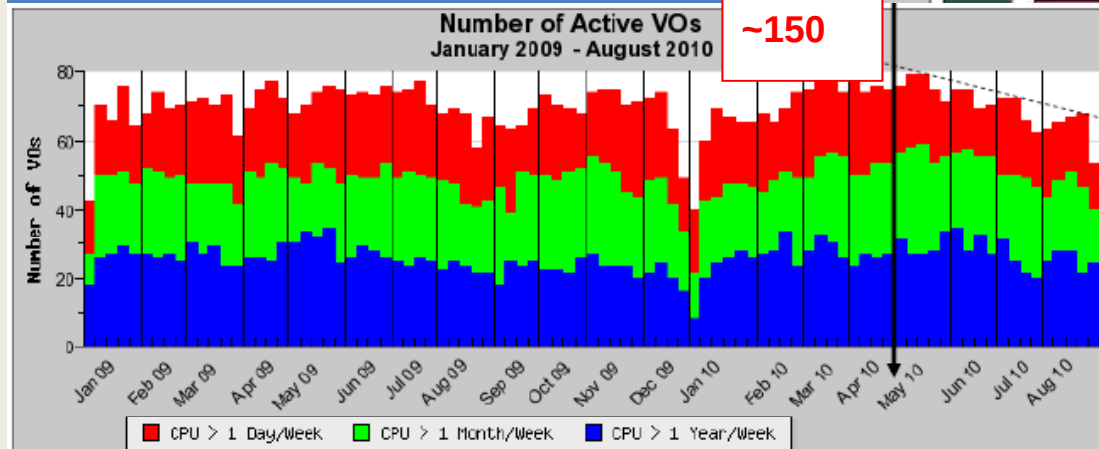
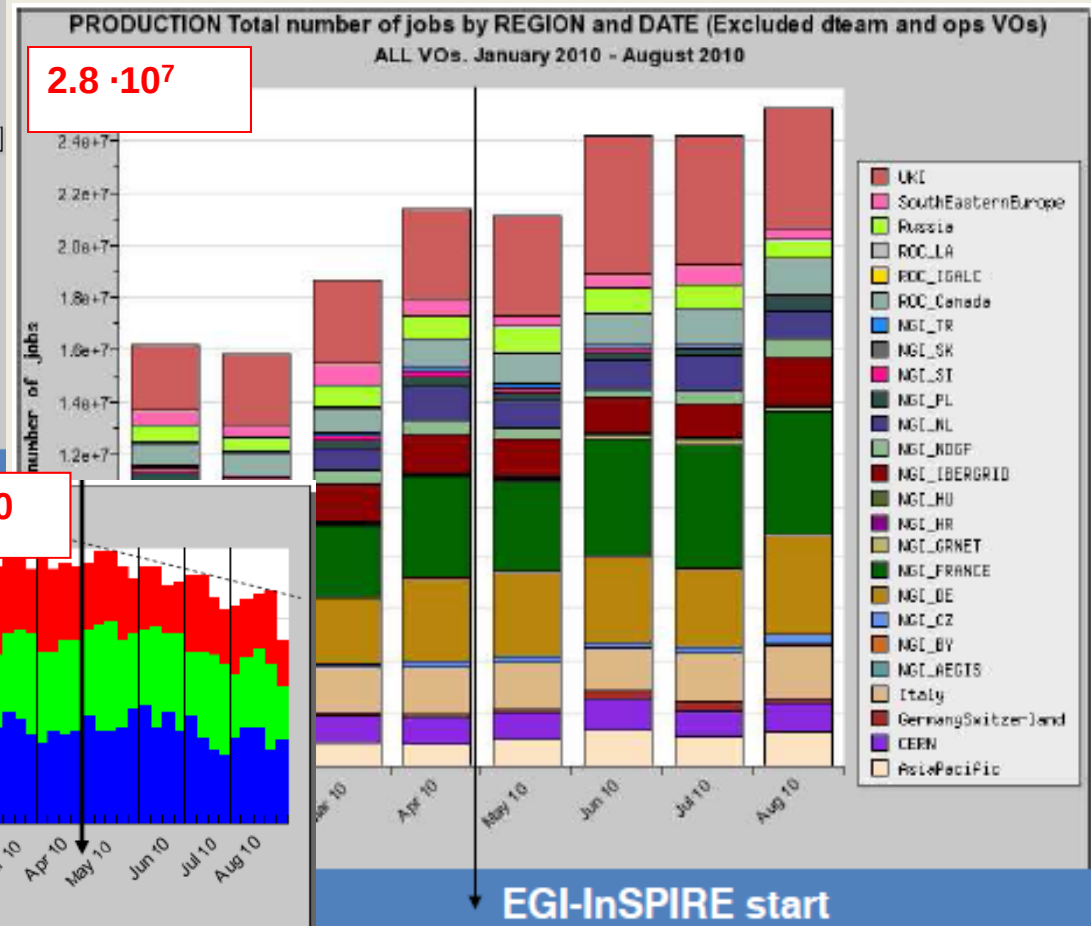
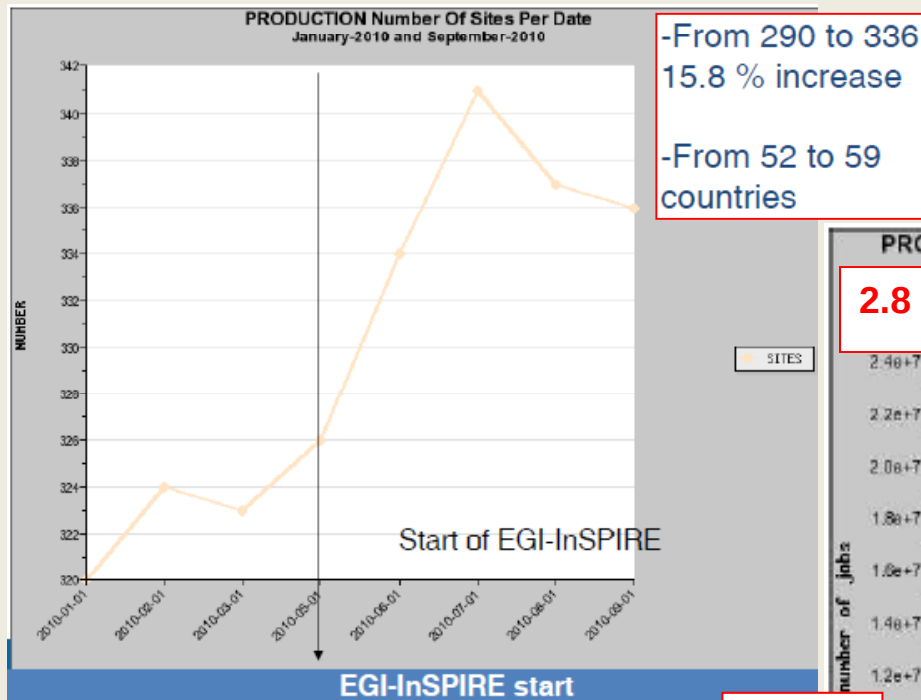
- Project cost € 72 M
- Total Effort ~ € 330 M
- Effort: 9261 PMs

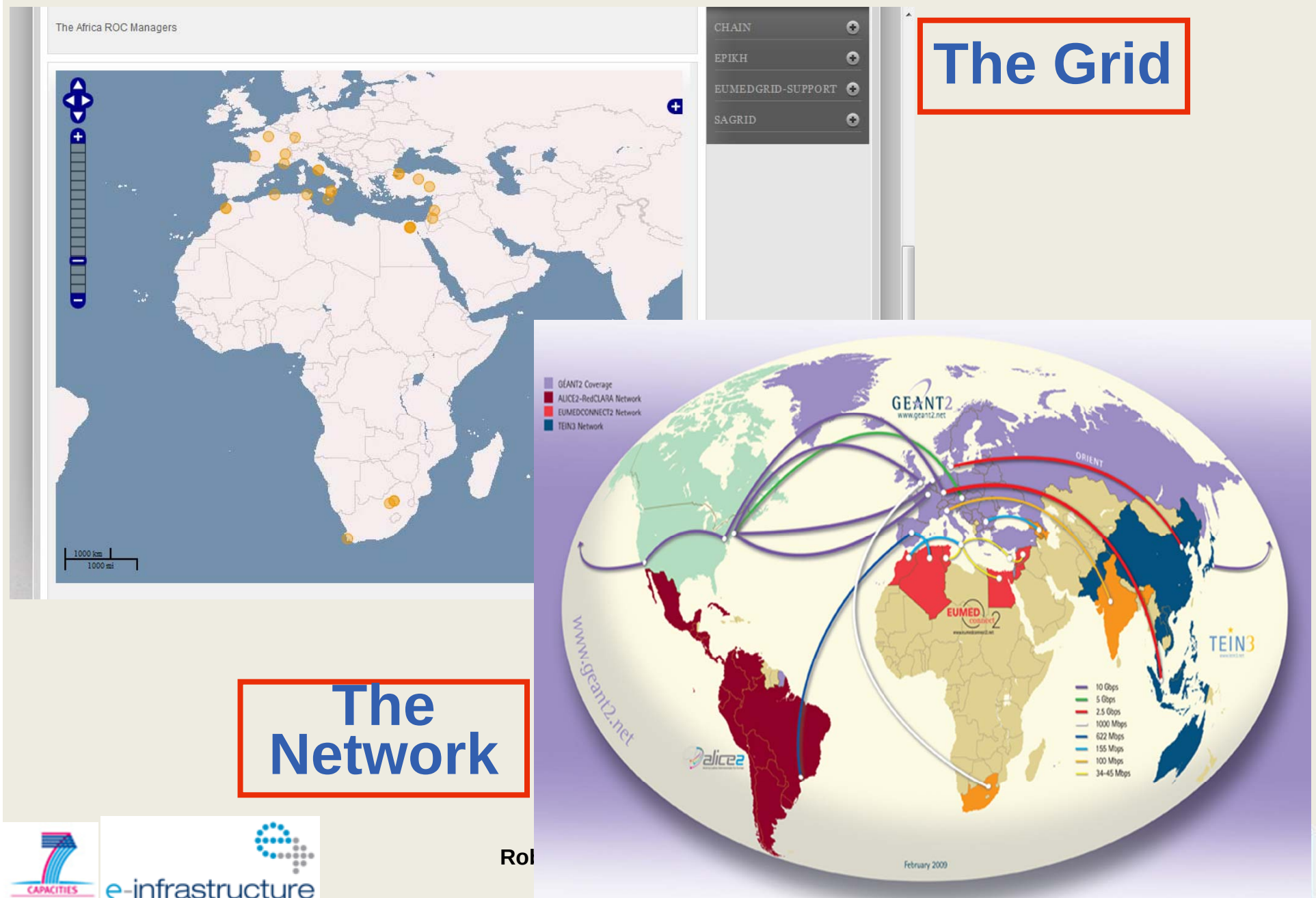
Project Partners (51)

- EGI.eu, 40 NGIs,
- 2 EU Int. Res. Orgs.
- Asia Pacific (8 partners)



- **Continue towards a sustainable production infrastructure:**
 - With infrastructure providers in Europe and around the world;
 - With new DCI technologies as they mature;
- **Provide support to current structured international research communities:**
 - Sustain current domain specific services;
 - Attract new user communities (e.g. ESFRI).



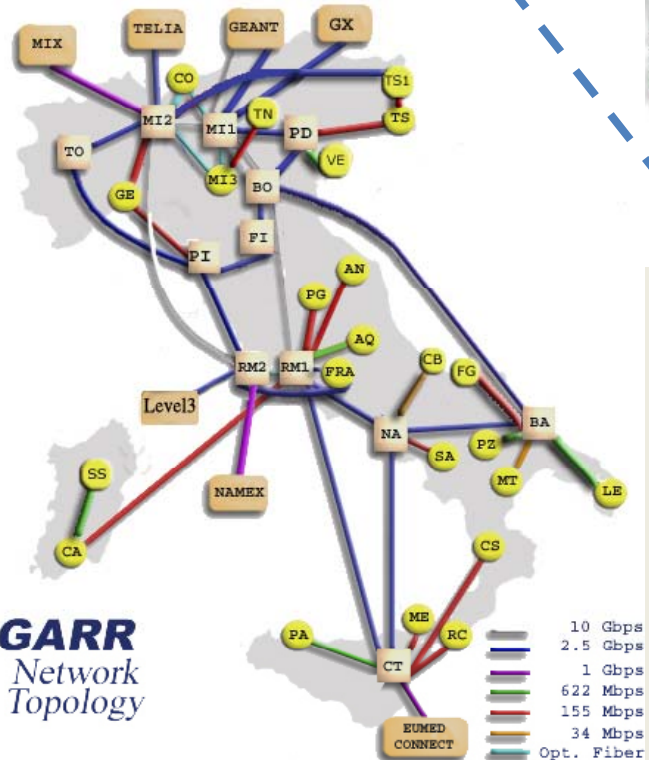


The Italian e-Infrastructure

(interoperable with gLite m/w as common communication protocol)



The Network



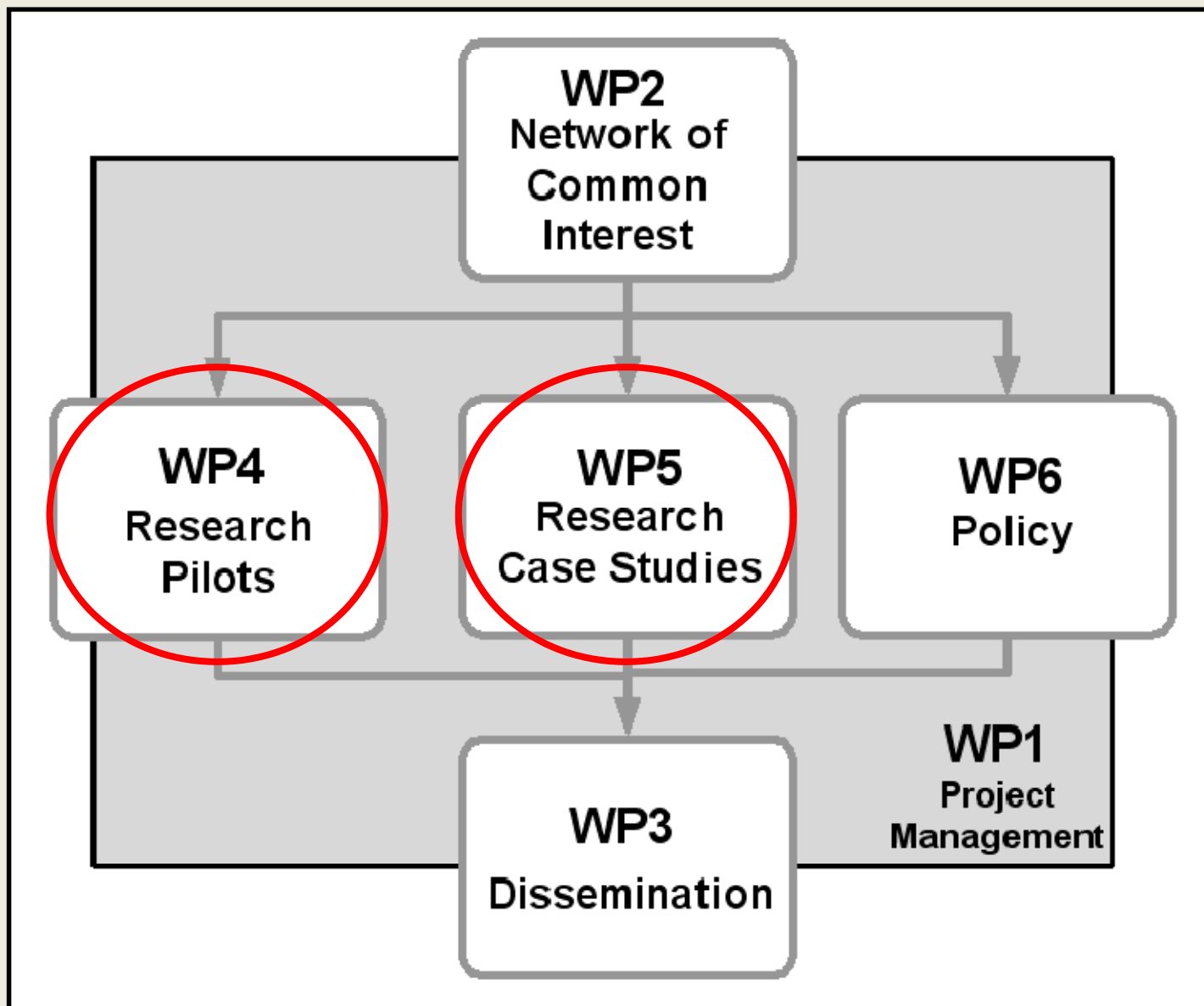
The Grid



Roberto Barbera, COMETA
Rome, 24/01/2011

- Full title: International Network for a Digital Cultural Heritage e-Infrastructure
- Support Action co-funded by the European Commission under the call FP7-INFRASTRUCTURES-2010-2
- Grant Agreement no.: 261324
- Duration: 24 months
- Start: 1st of September 2010
- Budget (Total): ~682 k€
- Budget (EC contribution): ~600 k€ (~88%)
- Person months: 113
- Project Manager: Rossella Caffo (ICCU)

- To establish a long-term network of common interest, following a tried and tested approach using working groups and high-quality publications;
- To deliver a range of demonstrators (two pilot applications and three case studies) that address the main priorities for the research in the Digital Cultural Heritage (in each case, these are not new R&D but instead the porting of existing work to the e-Infrastructure platform, with subsequent analysis and lesson learning);
- To establish the dialog between e-Infrastructure and DCH experts that are not used to work together.



- Pilot applications (WP4):

- E-Infrastructure-enabled semantic search for cultural repositories, where data aggregated for the MICHAEL project are being semantically enriched and, respective search, retrieval and publishing approaches are evaluated for porting to the e-Infrastructures platform provided by the EUMEDGRID e-Infrastructure and EUMEDCONNECT2 network.
- E-Infrastructure-enabled e-Collaborative Digital Archive protected by access control and rights management. This pilot will put in place two e-collaborative repositories: the Architectural and Archaeological Heritage present in Mediterranean Area (MED Repo) and the Federico De Roberto literary works archive.

- Research case studies (WP5):

- Digital preservation using e-Infrastructures;
- Virtual exhibitions using e-Infrastructures;
- Geocoded digital cultural content using e-Infrastructures.

No	Name	Short name	Country
1	Istituto Centrale per il Catalogo Unico delle biblioteche italiane e per le informazioni bibliografiche	ICCU	Italy
2	CONSORTIUM GARR	GARR	Italy
3	COMETA CONSORZIO MULTI ENTE PER LA PROMOZIONE E L'ADOZIONE DI TECNOLOGIE DI CALCOLO AVANZATO	COMETA	Italy
4	MINISTERE DE LA CULTURE ET DE LA COMMUNICATION	MCC	France
5	NATIONAL TECHNICAL UNIVERSITY OF ATHENS	NTUA	Greece
6	FUNDACIO PRIVADA I2CAT, INTERNET I INNOVACIO DIGITAL A CATALUNYA	I2CAT	Spain
7	CENTER FOR DOCUMENTATION OF CULTURAL AND NATURAL HERITAGE CULTNAT	CULTNAT	Egypt
8	Department of Antiquities	DA	Jordan
9	Türkiye Cumhuriyeti Kültür ve Turizm Bakanlığı	SGB	Turkey
10	AAS AVTORSKA AGENCIJA ZA SLOVENIJO DOO	AAS	Slovenia

10 Partners from 8 Mediterranean countries

The COMETA Consortium (www.consorzio-cometa.it)

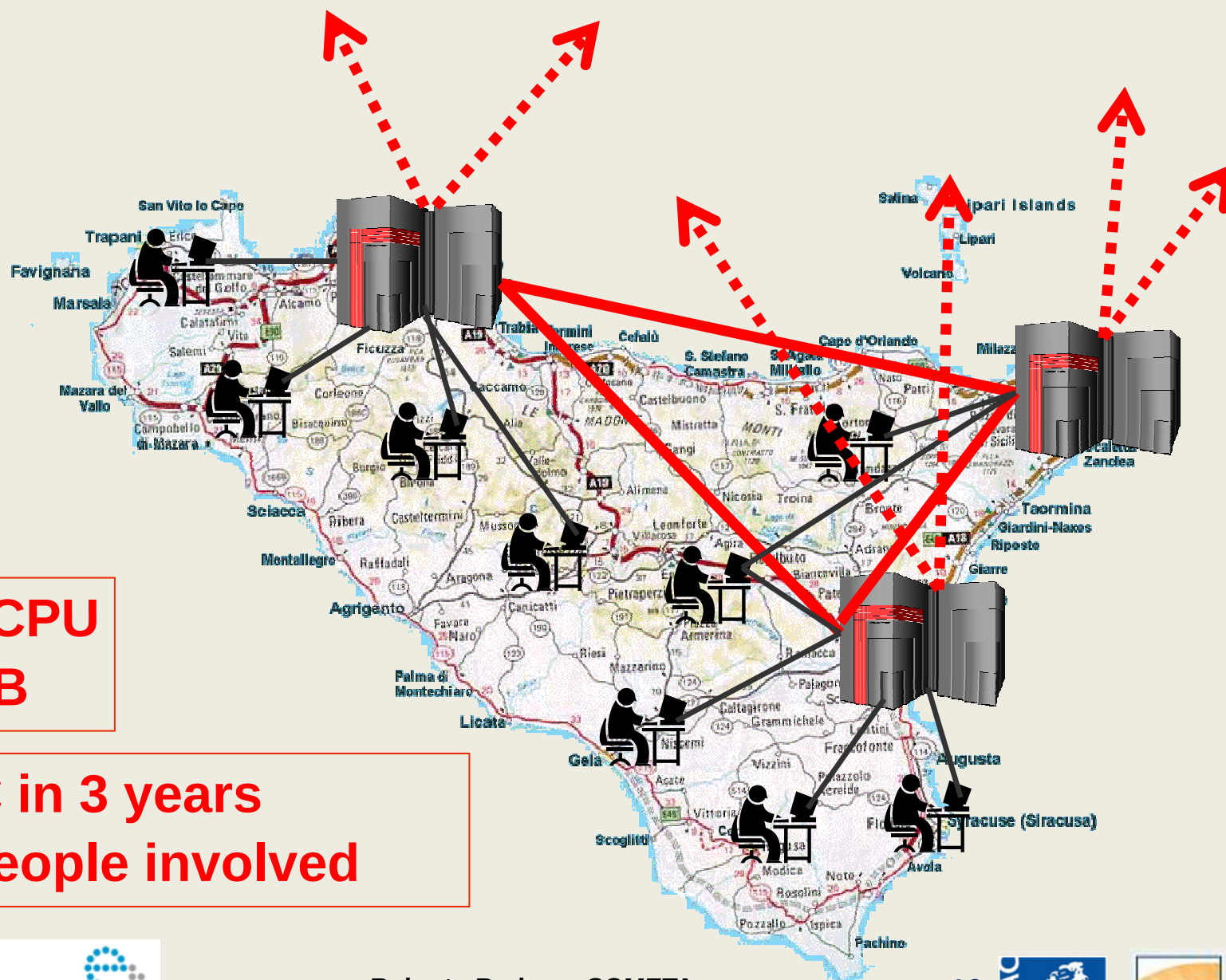




FESR



Ministero
dell'Università e della Ricerca



~2000 CPU

~250 TB

~12 M€ in 3 years

>270 people involved





1. ~2000 cores AMD Opteron 2218 rev. F
2. 2 GB of RAM per core
3. Commercial LRMS (LSF)
4. Infiniband-4X (for MPI applications)



1. ~250+ TB of storage
2. Distributed parallel filesystem (GPFS)

- A comprehensive database is available at:
 - www.pi2s2.it/applications (**122** applications developed)



- To know more, look at:
 - agenda.ct.infn.it/materialDisplay.py?materialId=12&confId=24
 - agenda.ct.infn.it/materialDisplay.py?materialId=20&confId=87

- High performance storage systems:
 - Geographically distributed replicas of files
 - Fault-tolerant digital preservation;
- Simplified authentication systems:
 - Single sign-on;
- Fine-grained authorization systems:
 - Allow to define exactly who:
 - Individual user(s), group(s), organization(s), the whole world;
 - And what:
 - Read/edit/delete data and metadata;
 - Search, browse;
 - Creation of new repositories;
 - Role assignments;
 - Etc.

- gLibrary is the INFN/COMETA **platform** that provides a simple yet powerful system to **store, organize, search** and **retrieve** “**digital assets**” in repositories built on e-Infrastructures
 - hides the underlying technical details to the users
- What we mean by “**digital asset**”:
 - data + metadata

gLibrary architecture

gLibrary uses many services
of the European Grid
middleware gLite

- Digital Asset of different kind (or type) are organized into **repositories**:
- “**types**” e “**collections**” definition by the administrator(s) of a given **repository**. Assets can be organized by:
 - **type**:
 - a list of attributes to describe a given kind of asset
 - hierarchical (a child type shares and extend parent’s attributes)
 - **filters**: subset of attributes queried during browsing
 - and/or **collection**:
 - group together related assets belonging to different **types**
 - useful to define subsets of assets of the same type (i.e. favorites, playlists)
 - the same asset can belong to multiple collection (sort-of tagging)

Collections

Repository

- Assets are browsed selecting a **type** (or a **collection**) and then one of its **filter**:
 - attributes of the selected types/collection, chosen from a defined list, used to narrow the result set
- Filter application is cascading: the selection of a filter value dynamically influences subsequent filter values (“à la iTunes” browsing)

- Users can **upload** their local assets to one or more (i.e., creating replicas) Storage Elements of the Grid;
- **Download** from Grid storage to user PC/laptop with a click:
 - choosing a replica from a list:
 - **future functionalities: storage maps, closer storage;**
- Transfers (handled over HTTPS) are authorized by the digital certificate imported into the user's browser:
 - only authorized users can download a given assets;
 - guest access is supported through **robot certificates**.

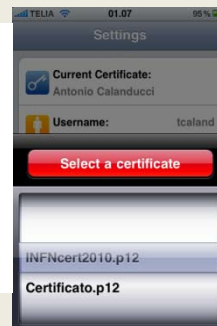


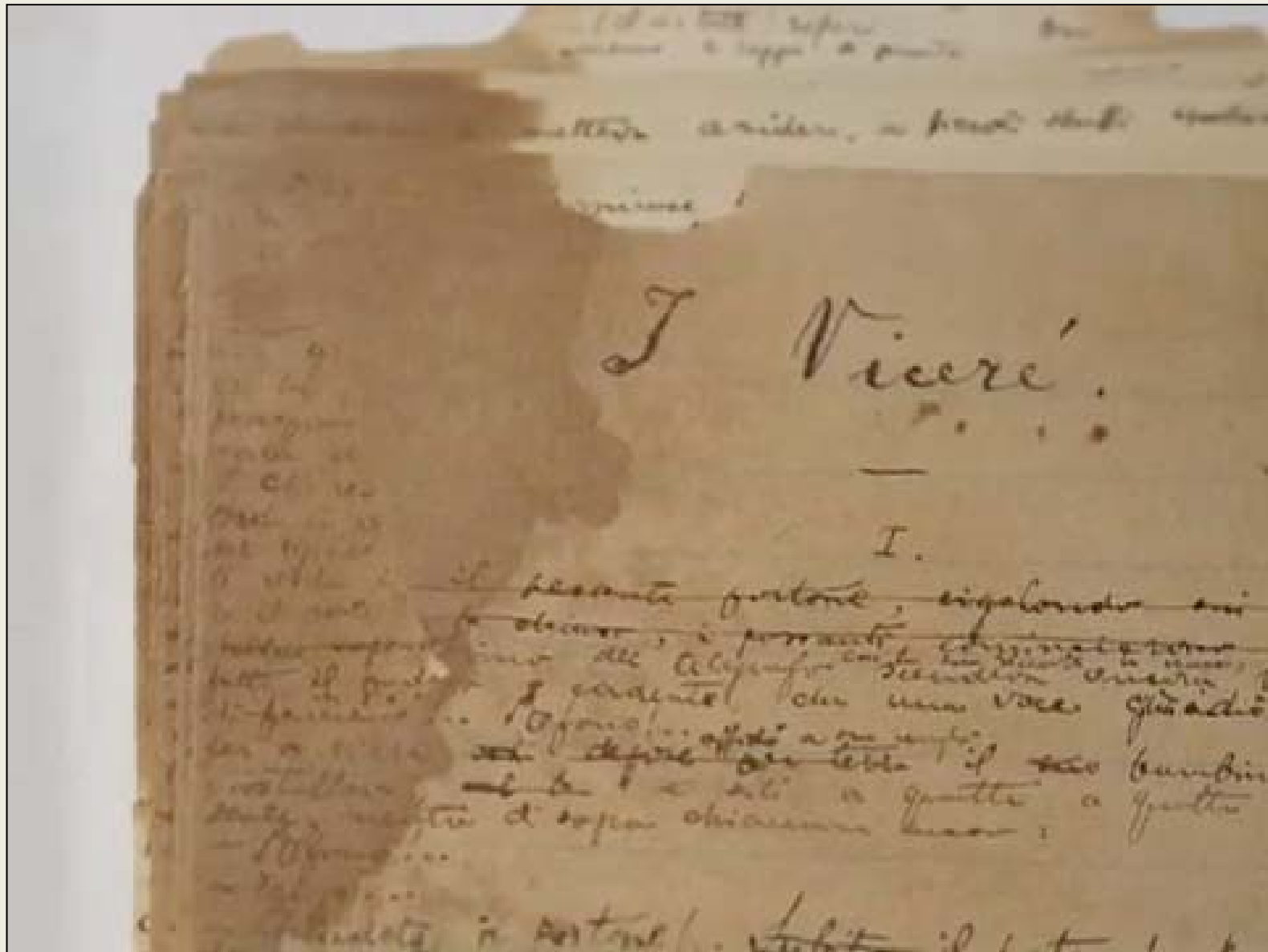
- Mobile interface to access Grid-based Digital Libraries:
 - *iPhone, iPod Touch, iPad;*
- Advantages:
 - provides an extremely **intuitive** and **touch-based** user interface to Grid storage elements and metadata, especially for non-expert users;
 - (to do) automatic selection of the closest replica, according to the user physical location retrieved by the integrated GPS;
 - offline access to the assets already downloaded.



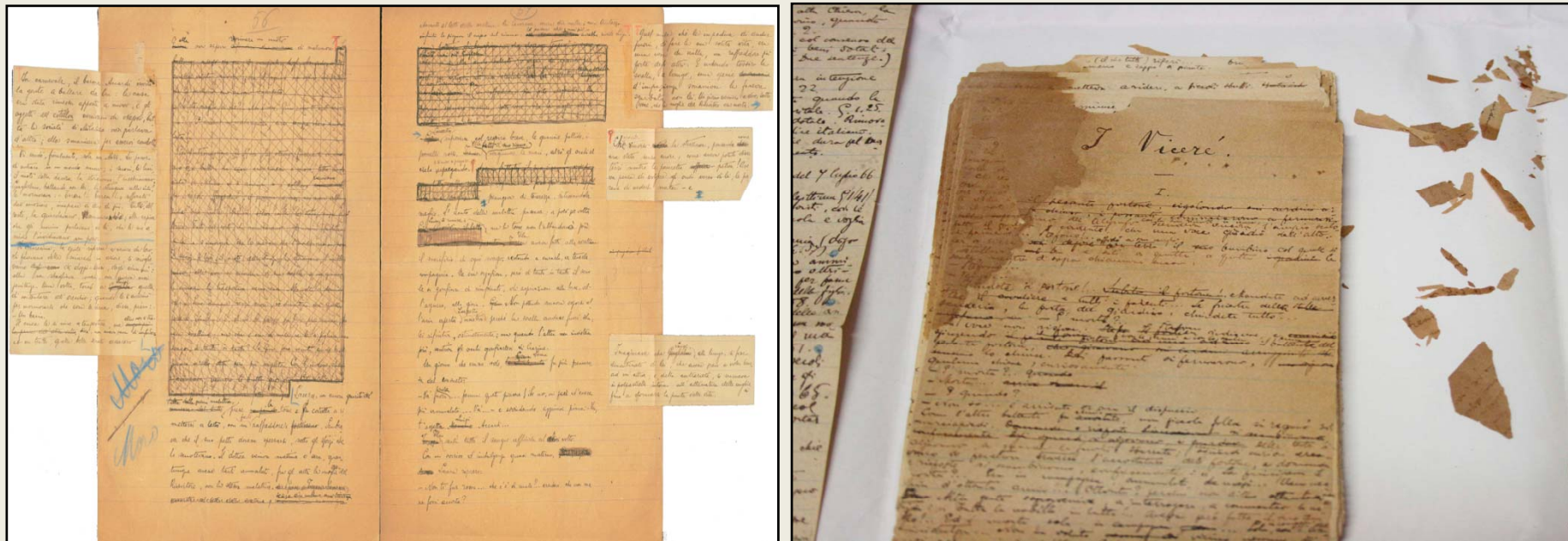


**Download and
select a X.509
Certificate**





- Some sheets are damaged (mold, crumbed pieces) and need physical restoration;
- Digitisation is mandatory to avoid the loss of this heritage; some works are still unpublished and relevant for the humanistic community.





- Digitalization of manuscripts, typescripts, printed works:
 - TIFF Files, one per page, 600 dpi, about 100 MB for A3
 - High resolution scans for in-depth examination;
 - PDF, one per work, 300 dpi, varying file sizes 40-400MB
 - Overall examination of works;
 - 8000 sheets/scans, 3 Terabyte of disk space;
 - Different physical formats, A3/A4/custom size;
- Embedded Metadata:
 - TIFF with embedded metadata to provide scan physical features and information about the content:
 - ImageWidth, ImageHeight, XResolution, FileSize, CreationDate, ModifyDate
 - Description, Keywords, CaptionWriter, Title, Author, Copyright Status, Copyright Notice;
 - Added with Photoshop after the digitisation phase (Adobe XMP format).

- Make manuscripts accessible to the humanistic community
 - Always on-line: 24 x 365
 - Available from everywhere
 - Simple and easy-to-use interface for non-expert people
- Quickly find the desired document
 - Document organization according the physical and semantic metadata
 - Organization by type/collections
 - Dynamic filtering of search result sets according the selection of one or more document metadata
- Long-term preservation (digital preservation)
 - Multiple copies (replicas) spread over different geographical sites
 - Reliability of storage systems and replica redundancy to achieve real preservation

- Goals:
 - store the 8000 scans of De Roberto Heritage
----> **Data Grid Storage Elements**
 - enable an ubiquitous and 24/24h access to scientists
---> **Web Application**
 - document organization for a quick search
---> **Metadata Services**
 - long-term digital preservation of data
---> redundancy through **Replicas** of files on several Storage Elements
 - simple and easy-to-use system for searches, organization, upload and download of digitalized documents on the Grid ----->

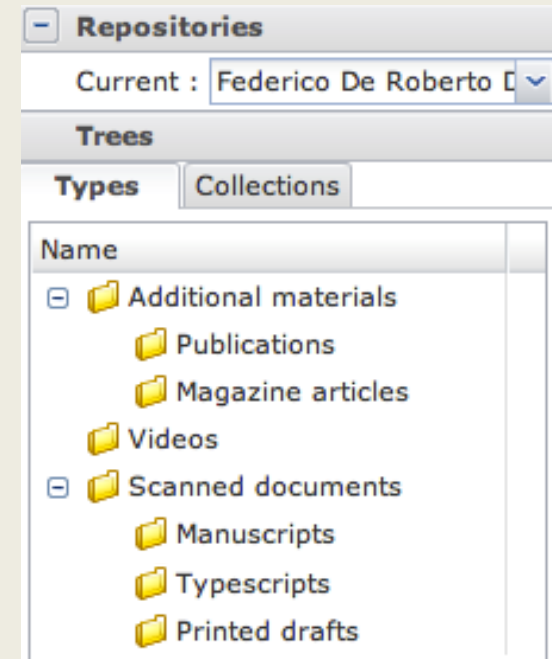


- Web 2.0 standards:
 - Javascript/AJAX/JSON on the client side;
 - PHP5 classes to implement business logic on the server side;
- Grid technologies:
 - Storage Element SRM interface to get the TURLs (Transfer URLs);
 - Transfers handled with GridFTP and X.509 cert auth HTTPS;
 - X.509 based Globus Security Infrastructure with the VOMS extensions to handle authentication and authorization (ACL based) on Metadata and Storage Elements;
 - All grid services implemented with the EGEE gLite middleware (DPM Storage Elements, AMGA Metadata Catalogue, LFC File Catalogue, VOMS Services);
- Other standards:
 - Subset of XMP Metadata:
 - Dublin Core Metadata set will be included in the next release:

- **Type** tree definition for the DR repository
- **Attributes** definition per type (e.g., manuscripts):



Attribute	Value
Title	la lupa
Author	federico de roberto, giovanni verga
Description	manoscritto della tragedia lirica ...
Keywords	verismo, federico de roberto, la lupa, ...
CaptionWriter	stefania iannizzotto, alessandro ...
CopyrightStatus	Copyrighted
PageNum	5
TotalPages	34
DocumentGenre	tragedia lirica
PublicationYear	1916
Publisher	officine tipo-litografiche barravecchia e balestrini
FileType	PDF
Resolution	300
ScanQuality	good



Filter defined per type.

Examples:

- DocumentGenre
- Title
- FileType
- ScanQuality
- DocumentType
- PublicationYear
- Publisher
- Location

The screenshot shows a web browser window with the URL http://glibrary.ct.infn.it/glibrary_new/browse.php. The page features a header with the 'DE ROBERTO digital repository' logo and navigation links: home, browse, search, upload, and logout. A search bar contains the text 'cultural heritage'.

On the left, a sidebar shows a tree structure under 'Trees' with 'Types' and 'Collections' tabs. The 'Collections' tab is active, showing a list of collections including 'Additional materials', 'Publications', 'Magazine articles', 'Videos', 'Scanned documents', 'Manuscripts', 'Typescripts', and 'Printed drafts'.

The main content area displays search results for 'la lupa' by Federico De Roberto. It includes a 'Filter attributes' section with dropdown menus for Title, FileType, Author, and ScanQuality. Below this, a table lists the search results:

Thumb	FileName	Title	Author	Description	PagNum	FileType	Docume
	002_la_lupa.tif	la lupa	federico de roberto	manoscritto della tragedia lirica in due atti, musicata da pierantonio tasca, pubblicata da officine tipo-litografiche barravecchia e balestrini, palermo 1919; altre edizioni: noto, tipografia rosario caruso, ottobre 1932; conservato presso la biblioteca di storia patria per la sicilia orientale	2	TIFF	tragedia
	003_la_lupa.pdf	la lupa	federico de roberto	manoscritto della tragedia lirica in due atti, musicata da pierantonio tasca, pubblicata da officine tipo-litografiche barravecchia e balestrini, palermo 1919; altre edizioni: noto, tipografia rosario caruso, ottobre 1932; conservato presso la biblioteca di storia patria per la sicilia orientale	3	PDF	tragedia
	003_la_lupa.tif	la lupa	federico de roberto	manoscritto della tragedia lirica in due atti, musicata da pierantonio tasca, pubblicata da officine tipo-litografiche barravecchia e balestrini, palermo 1919; altre edizioni: noto, tipografia rosario caruso, ottobre 1932; conservato presso la biblioteca di storia patria per la sicilia orientale	3	TIFF	tragedia

003_la_lupa.pdf

View Edit generic attributes Edit specific attributs Download Relations

Thumbnail



Details

FILE:	50
FileName varchar(255):	003_la_lupa.pdf
TypeID int:	2
CategoryIDs varchar(150):	
SubmissionDate timestamp:	2009-03-02 18:37:00
Description varchar:	manoscritto della tragedia lirica in due atti, musicata da pierantonio tasca, pubblicata da officine tipo-litografiche barravecchia e balestrini, palermo 1919; altre edizioni: noto, tipografia rosario caruso, ottobre 1932; conservato presso la biblioteca di storia patria per la sicilia orientale
Keywords varchar:	verismo, federico de roberto, la lupa, giovanni verga, officine tipo-litografiche barravecchia e balestrini palermo, tipografia rosario caruso noto, società di storia patria per la sicilia orientale catania, manoscritti letterari moderni, la.mu.s.a., facoltà di lettere e filosofia, università degli studi di catania, società di storia patria per la sicilia orientale
LastModificationDate	2007-02-17 19:53:00

003_la_lupa.pdf

View Edit generic attributes Edit specific attributs **Download** Relations

Replicas

https://inaf-se-01.ct.pi2s2.it/dpm/ct.pi2s2.it/home/cometa/deroberto/busta01/003_la_lupa.pdf

https://unipa-se-01.pa.pi2s2.it/dpm/pa.pi2s2.it/home/cometa/deroberto/busta01/003_la_lupa.pdf

https://unict-dmi-se-01.ct.pi2s2.it/dpm/ct.pi2s2.it/home/cometa/deroberto/busta01/003_la_lupa.pdf

More s



Roberto Barbera, CC-BY
Rome, 24/01/2011

- COMETA and DC-NET:
 - COMETA has signed a Cooperation Agreement with DC-NET;



- COMETA and INDICATE:
 - Develop an e-Culture Science Gateway (eCSG) with general user authentication and authorisation schemas;
 - Make two Digital Cultural Heritage pilot use cases available through the eCSG:
 - Digital Repository of Federico De Roberto works (De Roberto DR);
 - Digital Repository of the Architectural and Archaeological Heritage in the Mediterranean Area (MED Repo).

INDICATE The INDICATE Virtual Organisation

voms admin for VO: vo.indicate-project.eu Current user: Roberto Barbera
Register! VO management Subscriptions Configuration Other VOs on this server

Manage [Create a new user](#)

Users:

User	delete
Antonio Calanducci	user
Roberto Barbera	delete user
Robot	delete user
Robot	delete user

1-4 of 4

User details

User's DN & CA: /C=IT/O=INFN/OU=Personal Certificate/L=Catania/CN=Roberto Barbera
/C=IT/O=INFN/CN=INFN CA [delete this user](#)

User's common name: Roberto Barbera

User's email address: roberto.barbera@ct.infn.it

Membership details

Group name	Roles
/vo.indicate-project.eu	VO-Admin

User details

User's DN & CA: /C=IT/O=INFN/OU=Robot/L=COMETA/CN=Robot: Digital Repository of Federico De Roberto works - Antonio Calanducci
/C=IT/O=INFN/CN=INFN CA [delete this user](#)

User's common name: Robot: Digital Repository of Federico De Roberto works -

User's email address: antonio.calanducci@ct.infn.it

Membership details

Group name	Roles
/vo.indicate-project.eu	VO-Admin Assign role

7 CAPACITIES **e-infrastructure**

Roberto Barbera, COMETA
Rome, 24/01/2011

42 **MITBA** **ICCU**

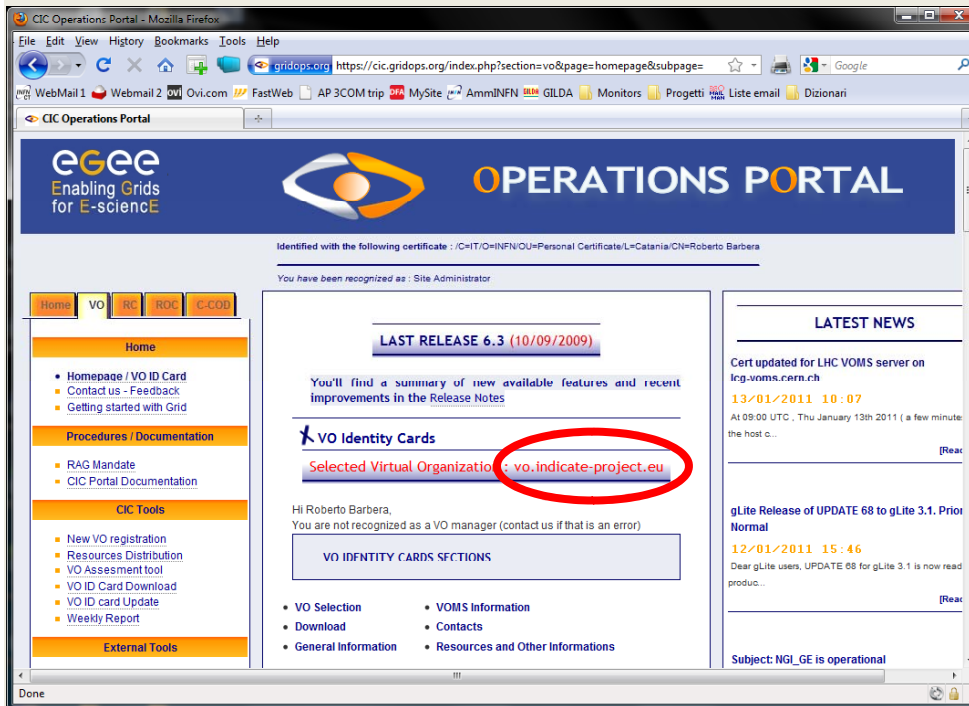
Robot certificates have been introduced to allow users, who do not have/want personal certificates and do not belong to any Virtual Organisation, to use the Grid; they are usually issued on smart-cards to be plugged on the machine where the service(s) is(are) executed.



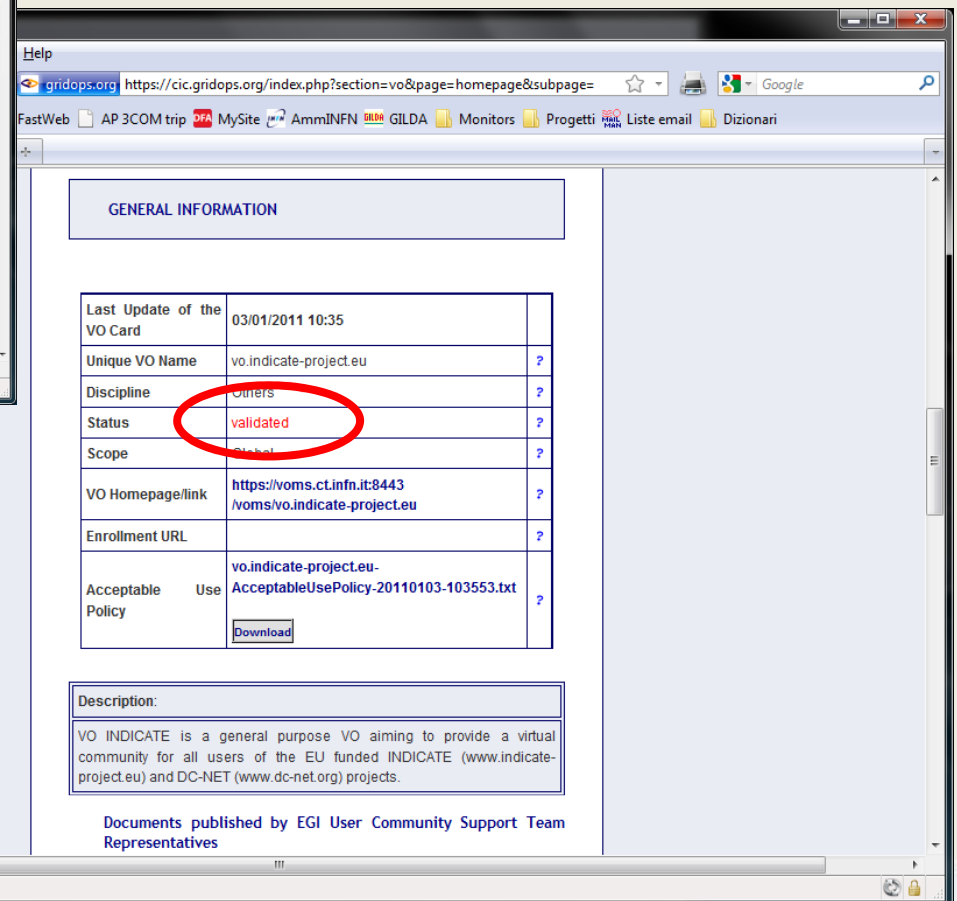
Basically, these certificates can be used to identify a person responsible for an unattended service or process acting as client and/or server for a Virtual Research community.

Function	Subscriber Name
Your identity: /C=IT/O=GILDA/OU=Robots/L=INFN Catania/CN=Robot:MrBayes - Giuseppe La Rocca	
Creating temporary proxy	
..... Done	
Contacting voms.ct.infn.it:15001	
[/C=IT/O=INFN/OU=Host/L=Catania/CN=voms.ct.infn.it]	
"gilda" Done	
Creating proxy	
..... Done	
Your proxy is valid until Thu May 8 21:42:05 2008	

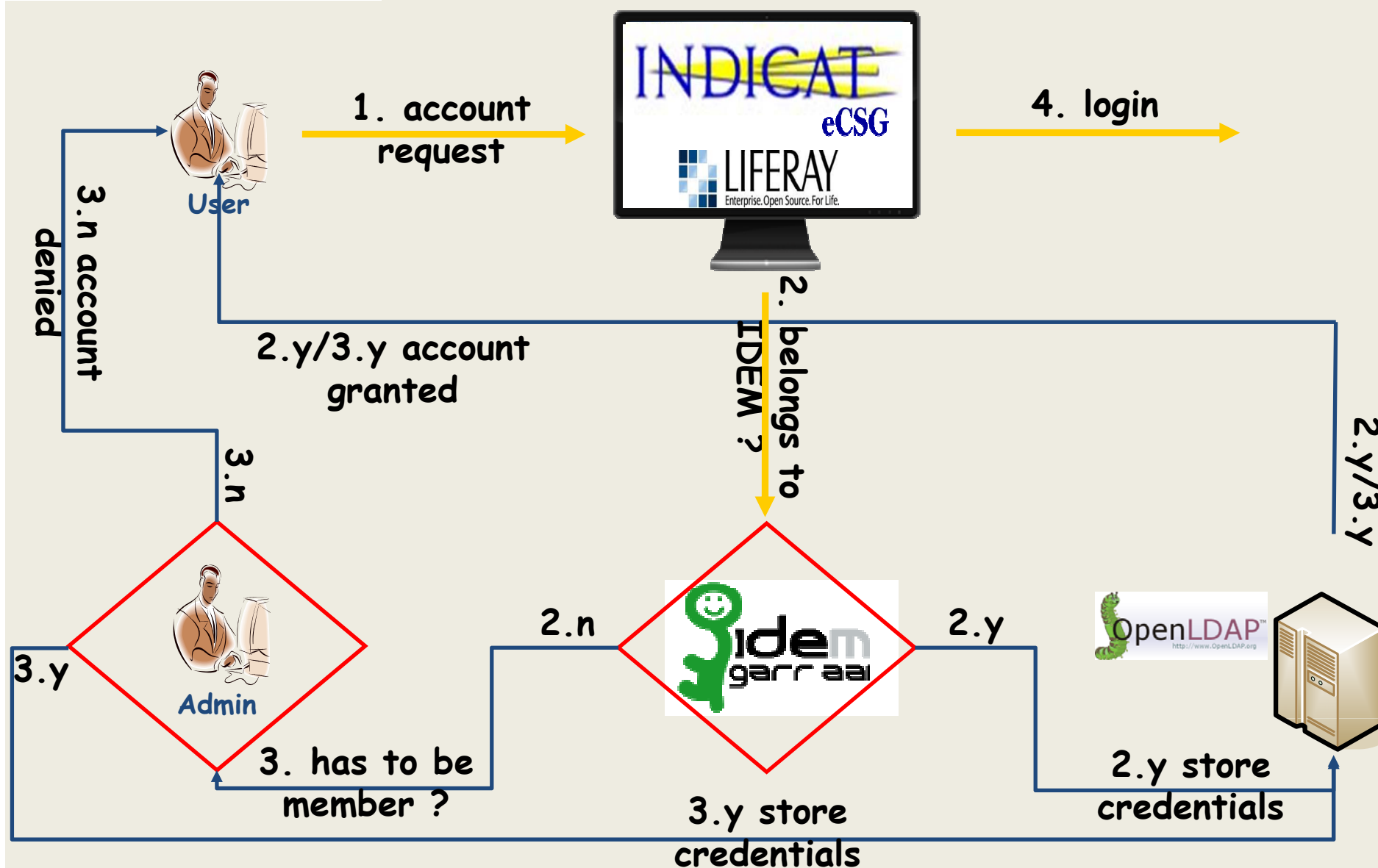
1) Now there is a «virtual home» for all DC-NET and INDICATE users and for all other people interested in DCH.



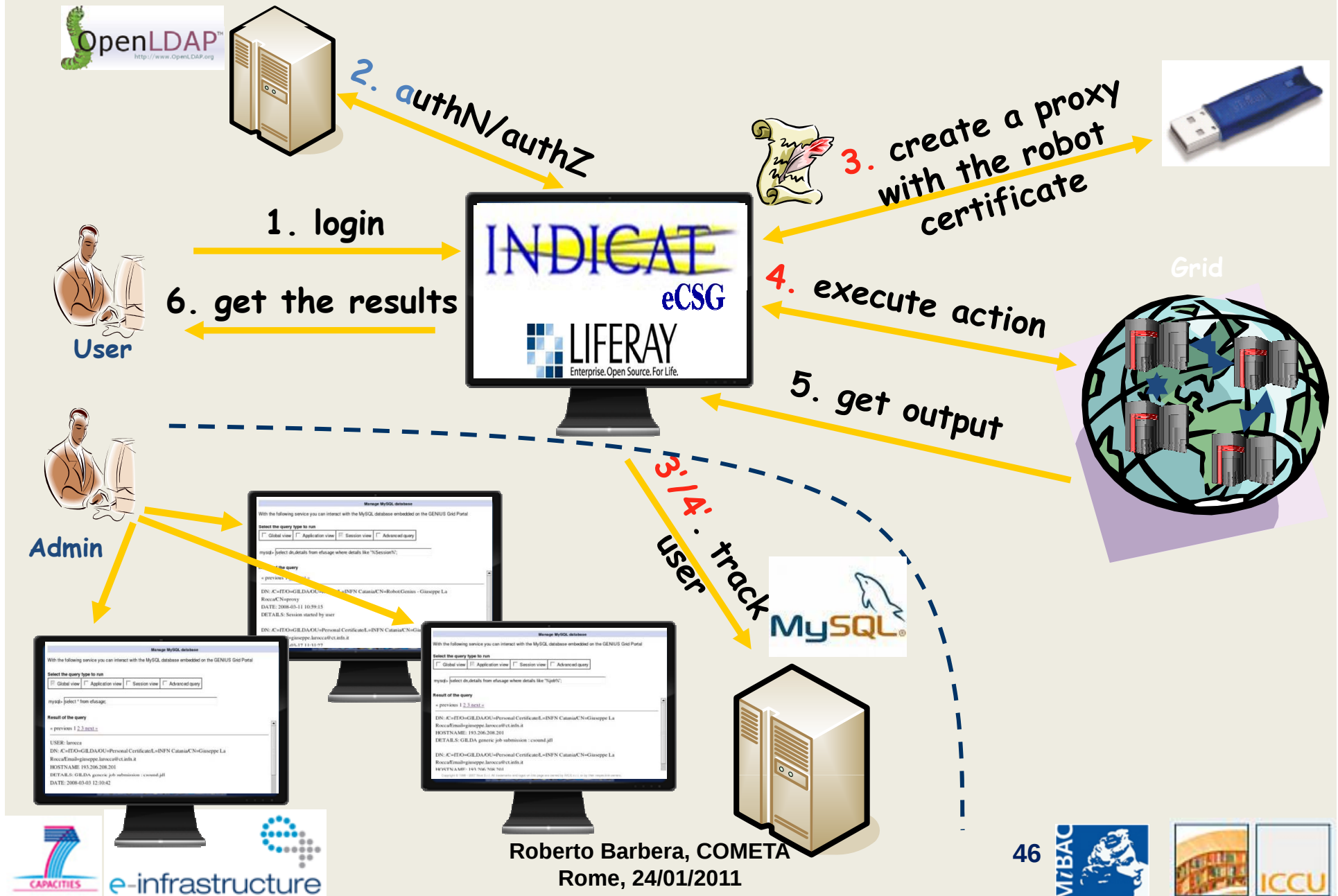
2) Now more NGIs/sites can contribute with their storage resources.



INDICAT The INDICATE e-Culture Science Gateway (1/5)



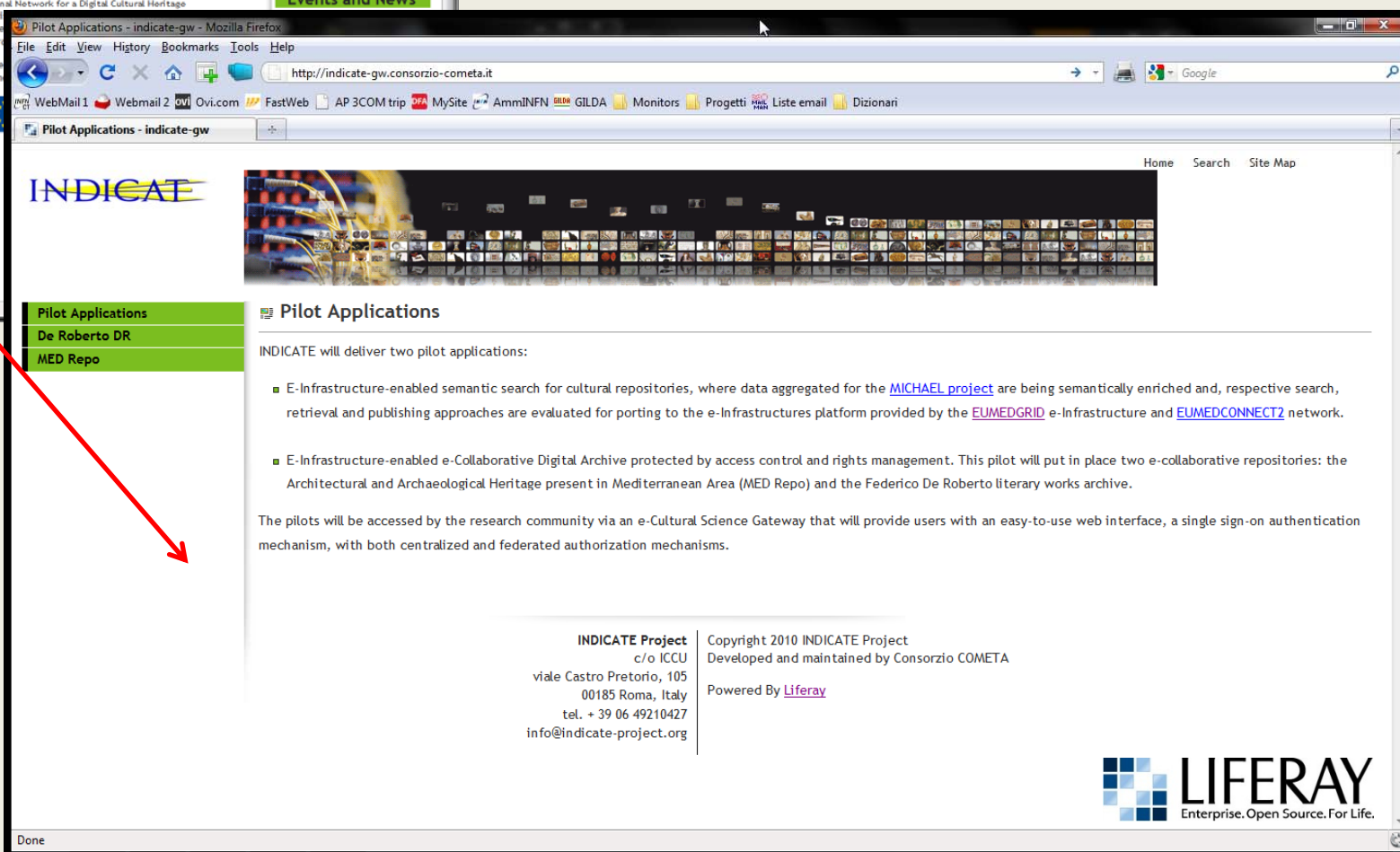
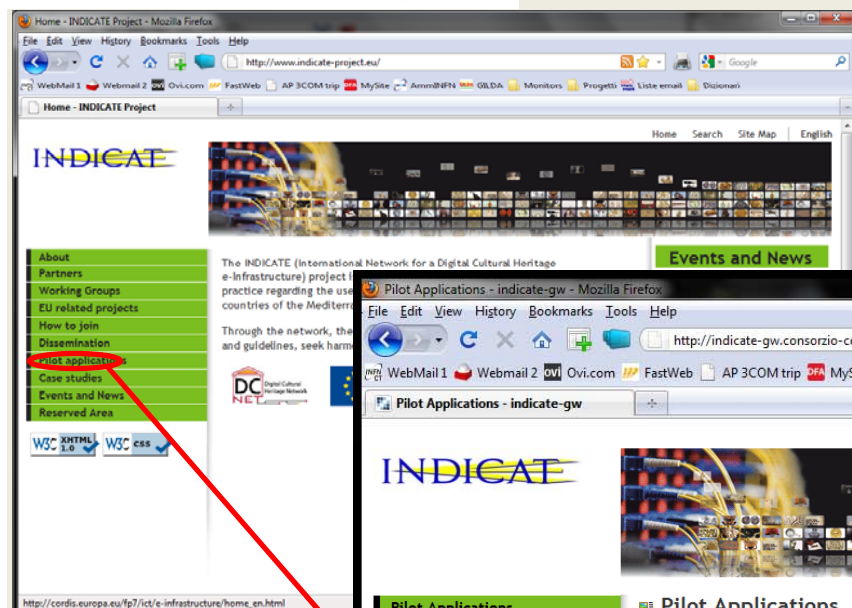
INDICAT The INDICATE e-Culture Science Gateway (2/5)



INDICATE The INDICATE e-Culture Science Gateway (3/5)

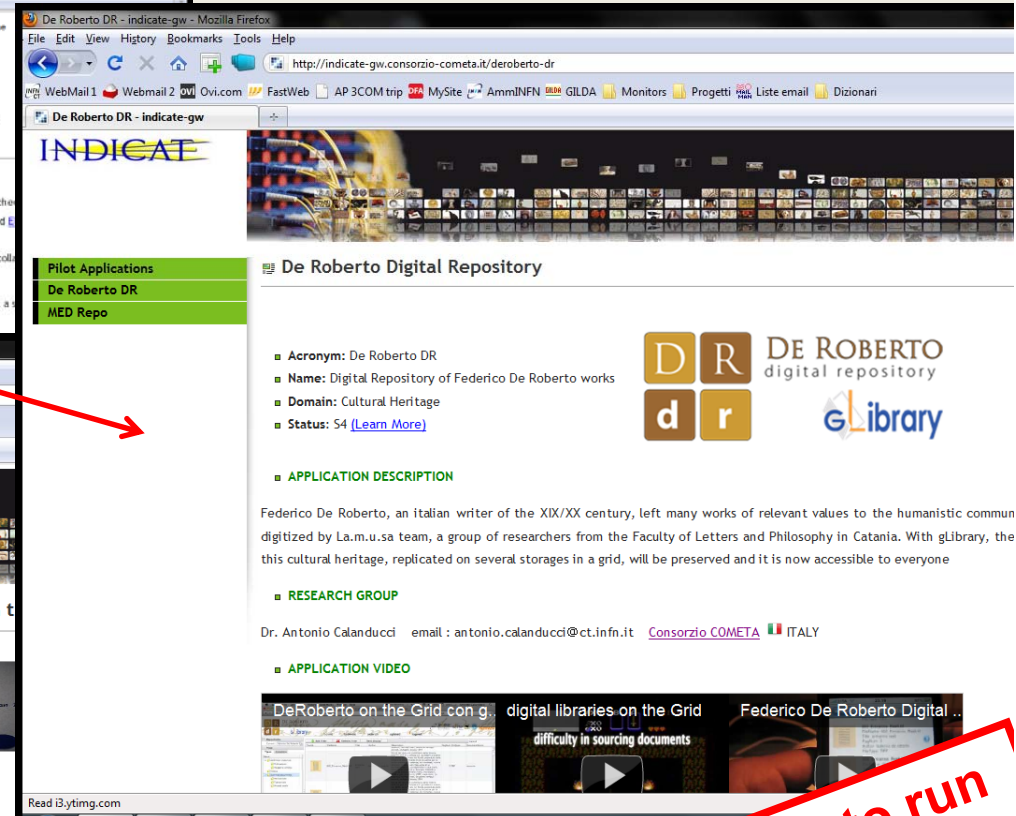
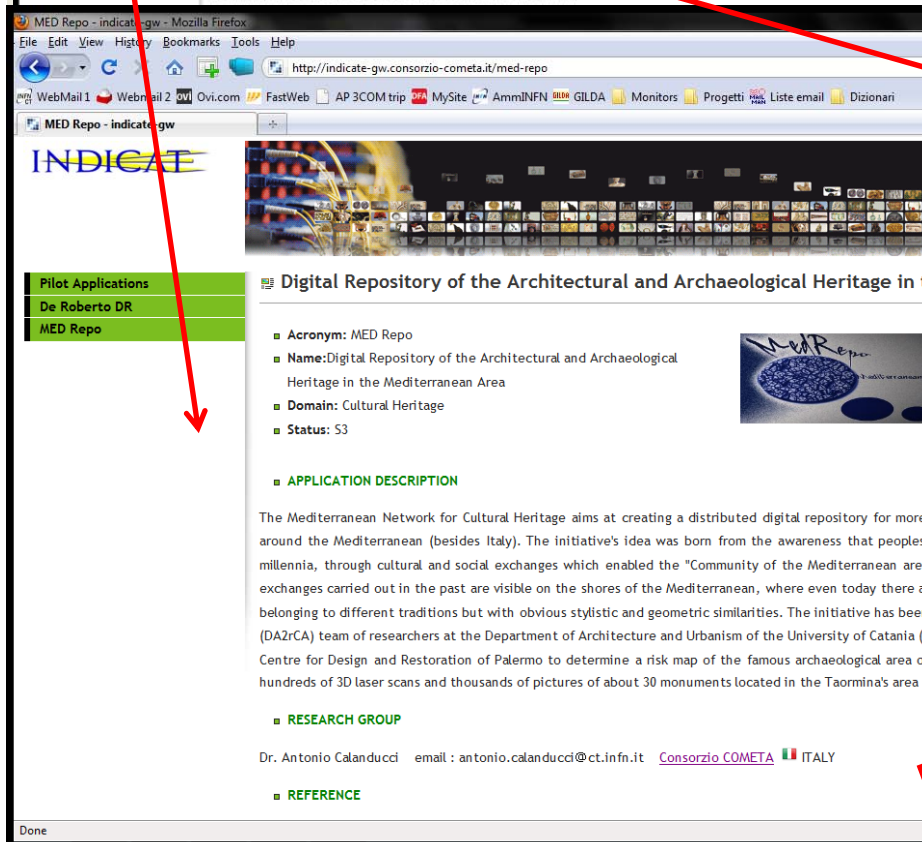
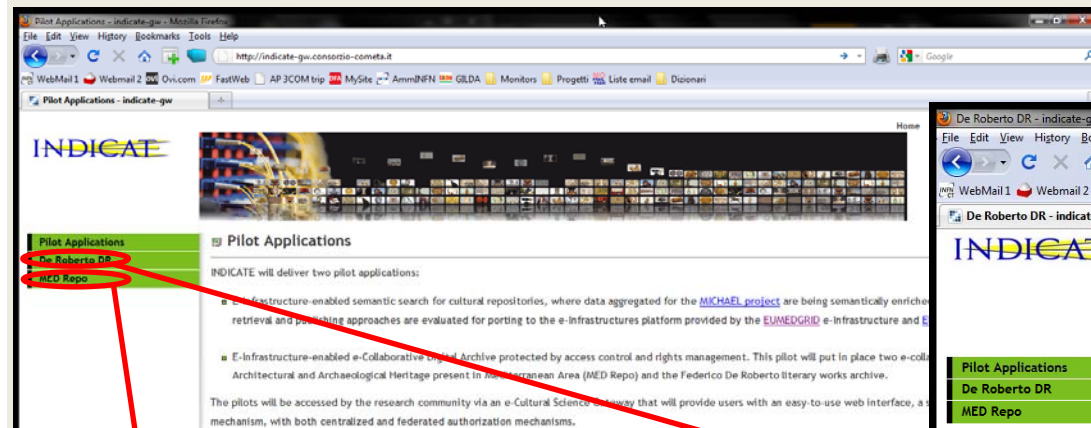
<http://www.indicate-project.eu>

<http://indicate-gw.consortio-cometa.it>



INDICATE The INDICATE e-Culture Science Gateway (4/5)

<http://indicate-gw.consortio-cometa.it>



There will soon be the possibility to run the applications on the Grid from within the eCSG

INDICATE The INDICATE e-Culture Science Gateway (5/5)

test-glibray - liferay-test

iceage-wn-14.ct.infn.it/test-glibray

Macity Facebook INFN WebMail Apple Yahoo! Google Maps YouTube Wikipedia News Popular javascript:isc.showL Other

LIFERAY
Enterprise. Open Source. For Life.

Welcome test-glibray

liferay-test test-glibray

gLibrary Test Portlet

This is the gLibrary Test Portlet portlet.

Thumb	FileName	Title	Author	Description	PagNum	FileType	DocumentGenre	DocumentType	Size
	004_la_lupa.tif	la lupa	federico de roberto	manoscritto della tragedia lirica in due atti, musicat...	4	TIFF	tragedia lirica	manoscritto	1...
	005_la_lupa.tif	la lupa	federico de roberto	manoscritto della tragedia lirica in due atti, musicat...	5	TIFF	tragedia lirica	manoscritto	1...
	001_la_lupa.tif	la lupa	federico de roberto, ...	manoscritto della tragedia lirica in due atti, musicat...	1	TIFF	tragedia lirica	manoscritto	1...
	010_la_lupa.tif	la lupa	federico de roberto	manoscritto della tragedia lirica in due atti, musicat...	10	TIFF	tragedia lirica	manoscritto	1...

Page 1 of 1

Displaying 1 - 50 of 50

**Federico De Roberto DR on Liferay.
Very preliminary!**

- INDICATE represents a concrete attempt to demonstrate the usefulness and effectiveness of e-Infrastructures for Digital Cultural Heritage;
- COMETA is one of the Grid resource providers of the Italian e-Infrastructure and, through it, of the European Grid Infrastructure;
- COMETA has a long experience in developing services for Virtual Research Communities, also in the field of Digital Cultural Heritage;
- The gLibrary framework is entirely based on gLite, one the most adopted middleware in the world;
- In the context of DC-NET and, mainly, INDICATE projects, COMETA will concretely show, through exemplar use cases, how Grid infrastructures can effectively be used by the Digital Cultural Heritage community in a simple and intuitive way.

- Proposal for an INDICATE Service Challenge:
 - Authorize the INDICATE Virtual Organisation in the Grid Storage Element of at least one Grid site of each project country;
 - Upload digitised contents of the Storage Elements and federate them through the definition and implementation of a common metadata schema (better if the contents have logical/cultural relations among them);
 - Create the digital repository with gLibrary with different user profiles (visitor, managers, creator, etc.) having different access privileges;
 - Ask the INDICATE end-users (and the general public, through a dedicated dissemination activity) to access the digital repository and provide feed-back;
- Time frame to prepare and run the Service Challenge: the next 6-8 months (in order to be able to report on its outcomes at the First Project Review);
- **Are any of the Italian Organisations cooperating with DC-NET willing to take part in the Service Challenge (as e-Infrastructure providers, data providers, or end users) ?**

Thank you for your attention

Roberto Barbera

roberto.barbera@ct.infn.it

www.indicate.eu

info@indicate-project.eu

- gLibrary contacts:
 - antonio.calanducci@ct.infn.it, glibrary@ct.infn.it
- Federico De Roberto repository:
 - <https://glibrary.ct.infn.it/>
- Videos:
 - <http://www.youtube.com/watch?v=VNN3OnpmUUU>
 - <http://www.youtube.com/watch?v=IhFFjHD8IsI>
- Publications:
 - A.Calanducci, R.Barbera, J.Sevilla, A. De Filippo, M.Saso, S. Iannizzotto, F. De Mattia, F.Vicinanza. “Data Grids for Conservation of Cultural Inheritance”, 1st International Workshop on Data Grids for e-Science (DaGreS09) at ACM International Conference on Computing Frontiers, May 18-20, 2009 (<http://www.computingfrontiers.org/2009/>)
 - <https://glibrary.ct.infn.it/m/DaGRes-editor.pdf>
 - A. Calanducci, C. Cherubino, L. N. Ciuffo, D. Scardaci, “A Digital Library Management System for the Grid”, Fourth International Workshop on Emerging Technologies for Next-generation GRID (ETNGRID 2007) at 16th IEEE International Workshops on Enabling Technologies: Infrastructures for Collaborative Enterprises (WETICE-2007), GET/INT Paris, France, June 18-20, 2007 (<http://etngrid.diit.unict.it/2007/index.html>).
 - https://glibrary.ct.infn.it/glibrary/downloads/gLibrary_paper_v2.pdf