



# Elixir, Lifewatch, Arts & Humanities

Giacinto DONVITO

INFN-Bari

# Outlook

- ▶ Stato delle collaborazioni
- ▶ Attività
- ▶ Servizi forniti e loro stato
- ▶ Sinergie tecnologiche con altri progetti
- ▶ Conclusioni

# Collaborazioni

4

## ► *Elixir:*

- INFN è membro della **JRU** di **Elixir-ITA**
- Come “**Tecnology provider**”, c’è anche: **GARR, CNRS, CINECA**
- La JRU Italiana è ora finalmente **membro ufficiale di Elixir**
- Abbiamo partecipato ai bandi per degli assegni di Ricerca banditi dalla JRU e ne abbiamo vinto uno:
  - Obiettivo: “Fare il porting di Galaxy su Mesos”
- Abbiamo un accordo per una “collaborazione di ricerca” con il CNR di Bari
- Collaborazioni anche in INDIGO
- Sedi che partecipano: per ora soprattutto Bari

# Collaborazioni

5

## ► *Lifewatch:*

- INFN è membro della **JRU** di **Lifewatch-ITA**
- Come “**Tecnology provider**”.
  - É l'unico partner prettamente tecnologico nella JRU
- La JRU Italiana è parte del consorzio europeo di Lifewatch sin dall'inizio.
- INFN è anche coinvolto nel Lifewatch Competence Center di EGI.
- Collaborazioni anche in INDIGO
- Stiamo esplorando collaborazioni possibili su bandi regionali
- Sedi coinvolte: Bari (Lecce)

# Collaborazioni

6

## ► *Arts & Humanities:*

- INFN è membro della **DARIAH-IT**
- La JRU Italiana è parte del consorzio europeo
- INFN è anche coinvolto nel DARIAH Competence Center di EGI.
- Collaborazioni anche in INDIGO
- Sedi coinvolte: Catania, Bari

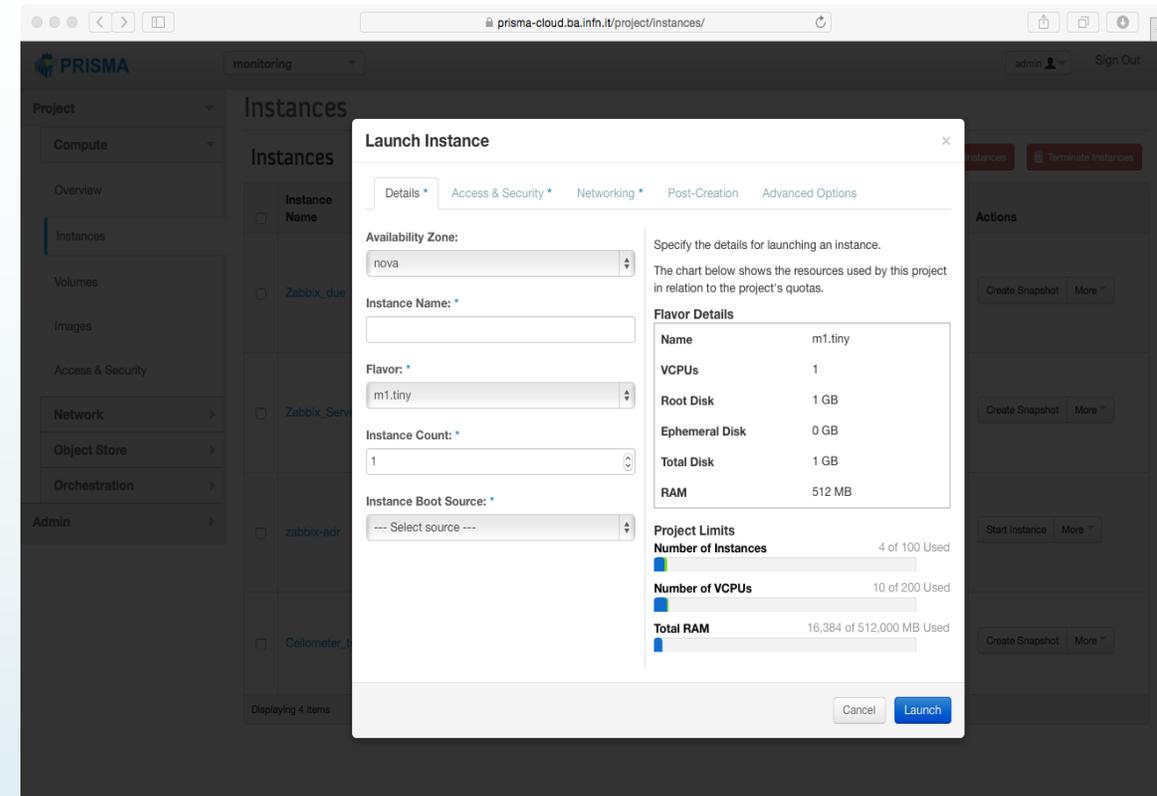
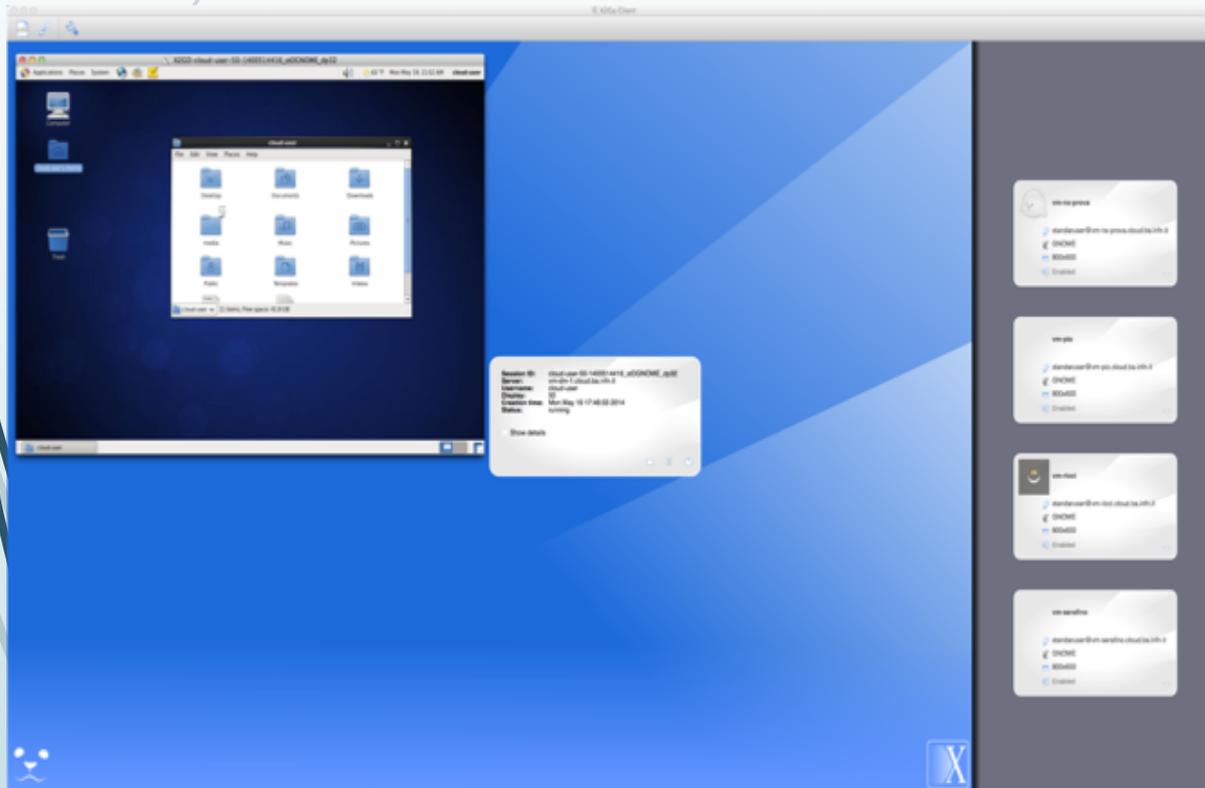
# Attività: Elixir/Lifewatch

7

- Supporto all'analisi dei dati da genomi
- Portali e complessi workflow
- Tecnologie usate:
  - Cloud computing e HTC
- Hosting dei dati critici
- SaaS per l'analisi dati
- Attività di sviluppo per l'implementazione di nuovi servizi

# Attività: Elixir/Lifewatch

8





# Attività: Elixir/Lifewatch

10



**ReCaS Science Gateway**

Benvvenuto sul portale ReCaS Science Gateway.

Dal menu "Applications" è possibile selezionare l'applicazione, fra quelle disponibili, da eseguire sulla nostra piattaforma di calcolo.

Nel menu "Help", selezionando l'applicazione di tuo interesse, troverai un pratico tutorial che ti illustrerà i semplici passi per poterli autenticare ed utilizzare l'applicazione.

Ricordati che è necessario effettuare il LOGIN prima di iniziare.

**PON ReCaS**

ReCaS è un progetto PON Ricerca e Competitività 2007-2013, Avviso 254/Ric. che si sviluppa all'interno delle seguenti circostanze favorevoli:

- capitalizzare i risultati della programmazione PON 2000-2006
- assecondare la sempre crescente necessità, da parte di numerosi settori della ricerca scientifica, di disporre di grandi risorse di calcolo e storage in contesti di calcolo distribuito.

Il progetto si compone di due parti: una di potenziamento infrastrutturale e l'altra di formazione.

Per maggiori informazioni ecco i nostri link: facebook, twitter e del PON.



## Msa PAD

Msa-PAD application is a multiple DNA sequence alignment framework designed to align conserved protein coding DNA sequences. The application accounts for either single or multiple protein domains coding sequences and uses this information in assembling its output. It is mainly useful for comparative genomics by the possibility to align genomes having different genic organization (i.e. bacterial genomes). It takes also into account and/or aligning genes' exons, including those undergone intron loss or gain, respecting genomic organization. BR /> Msa-PAD has two different alignment modes: (i) genome and (ii) gene. The difference between the two modes resides in the organization of the final alignment.

Genome Mode alignment, similar to super-gene alignment, keeps only fragment of sequences coding for unique protein domains and the final output is simply the concatenation of those fragments by randomly organizing the domains.

Gene Mode alignment respects the genomic organization of the input sequences; it identifies the most frequent domains order pattern. Consequently, domains order follows the increasing sites position of input DNA sequences.

Both modes share the following six steps:

1. It translates DNA sequences using a user-defined genetic code and frame/s by executing a custom Python script.
2. It makes use of PFAM-A [1] profiles information to assign translated sequences to a known conserved protein domain. This is obtained by searching against PFAM-A database using hmsearch (HMMer3.0 package) [2].
3. It elaborates protein sequence assignment taking into consideration frameshifts and intron gain or loss.
4. It groups sequences belonging to the same protein domain and consequently align them against the same domain using hmalign (HMMer3.0 package)
5. It Back-translates the protein alignments into DNA alignments.

## Execute our tool

Help  
 Upload file  
 Upload DNA sequence file in FASTA or .ZIP format:

Upload File

### Upload status

Current state Idle  
 File name  
 Status:

Reset list files

DNA sequence path file in FASTA or .ZIP format:

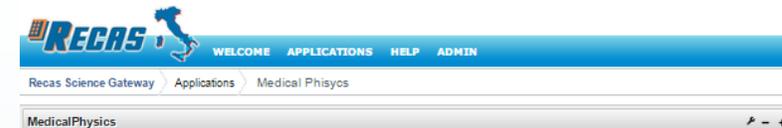
Alignment mode: \*  
 Genome  
 Genetic code: \*  
 1  
 Reading Frame: \*  
 1

Mail recipient

Execute

## Check status of your executions

Mail recipient



## Medical Physics

The Bari Medical Physics group offers a series of fully automated algorithms for the analysis of structural T1 weighted brain magnetic resonance scans. In particular the proposed workflows perform brain extraction and hippocampal segmentation.

More workflows will be available, stay tuned!

1. **Brain extraction:** This tool performs an automated brain extraction and calculation of intracranial volumes.
2. **Segmentation:** This method is based on active learning strategies for multi-atlas selection and Random Forest voxel-based classification. The user can pick up the number of templates to be used or just use the default settings (suggested)



## Execute our tool

Help  
 Upload file

UPLOAD IMAGE FILE (Only .nii or zip image)

Upload file  
 Nessun file selezionato   
  
 File uploaded:

Image file path:

Tool: \*  
 Brain\_extraction  
 Number of correlated images:  
  
 Mail recipient:

## Check status of your executions

Mail recipient:

Show

Results:

| Name         | Execution status |
|--------------|------------------|
| FisicaMedica | Done             |

Number of executions 4

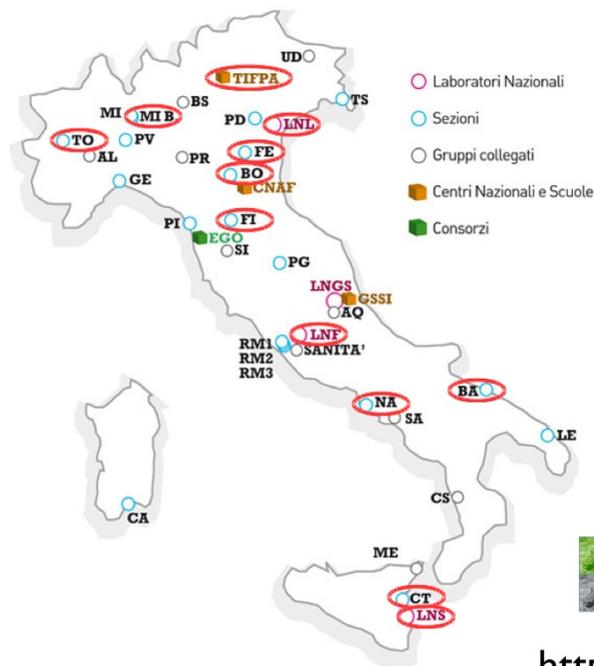
# Attività: Dariah

11

- ▶ Portali e complessi workflow
  - ▶ Sia in Cloud che in Grid
- ▶ File Catalogue and browsing
  - ▶ Sia in Cloud che in Grid

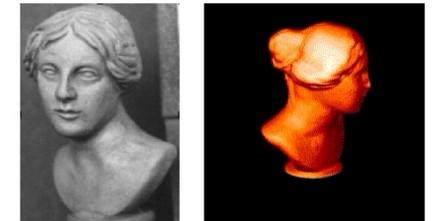
# Attività: Dariah

Offline data intensive conceptualisation  
(and contextualisation) in CH → INFN CHNet



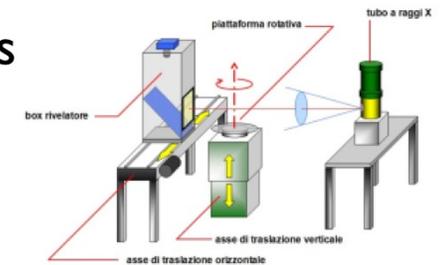
## ▶ *In situ* analyses

- ▶ Raman spectr.py
- ▶ XRD
- ▶ XRF
- ▶ X-tomography



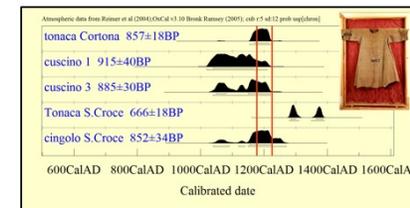
## ▶ Laboratory analyses

- ▶ PIXE, PIGE, BS, IBIL
- ▶ SEM
- ▶ k-edge radiography
- ▶ X-tomography



## ▶ Dating

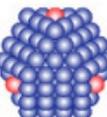
- ▶  $^{14}\text{C}$
- ▶ Thermoluminescence



## ▶ Collaborations:



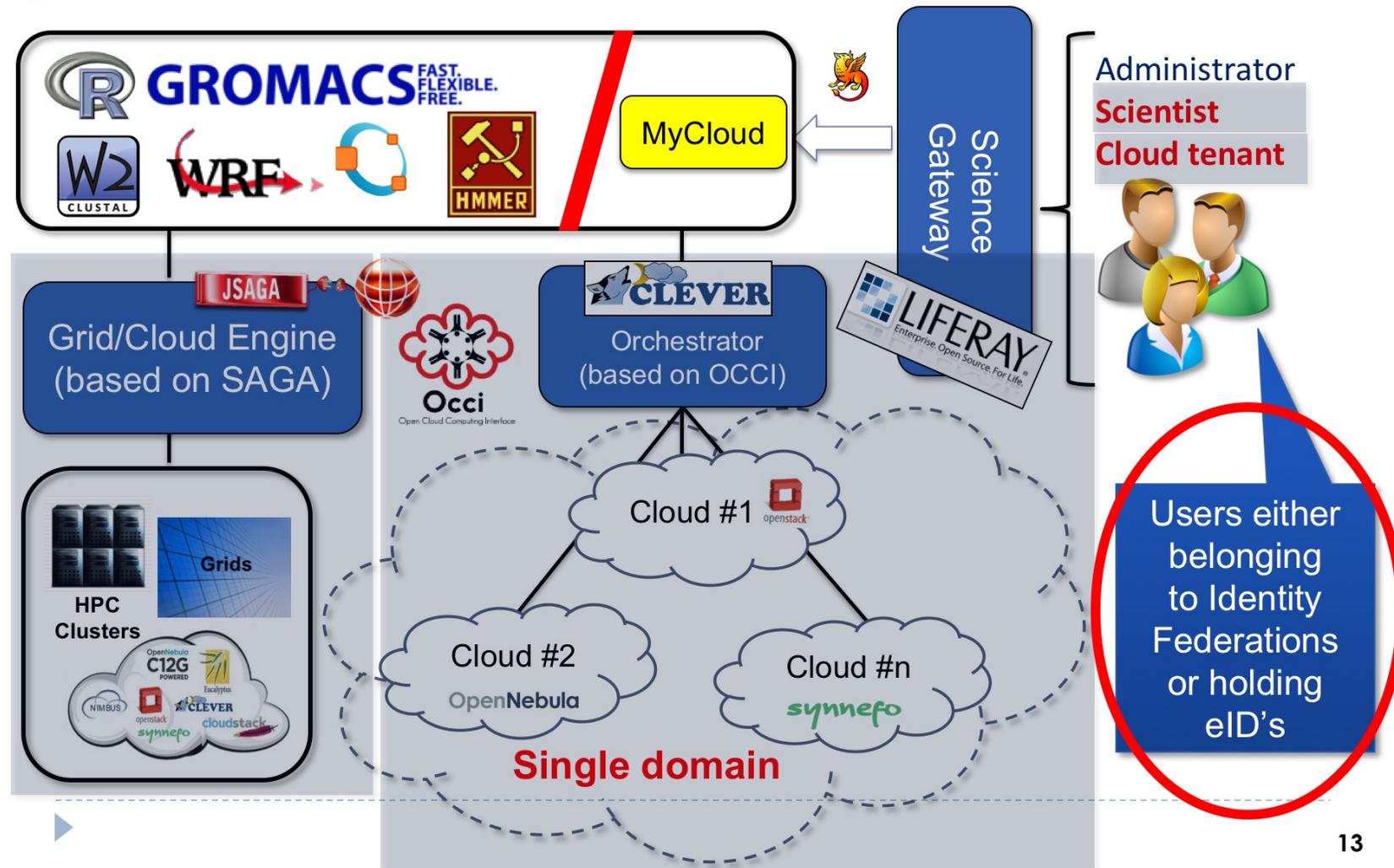
IPERION CH.it



# Attività: Dariah

13

## The Catania Science Gateway Framework



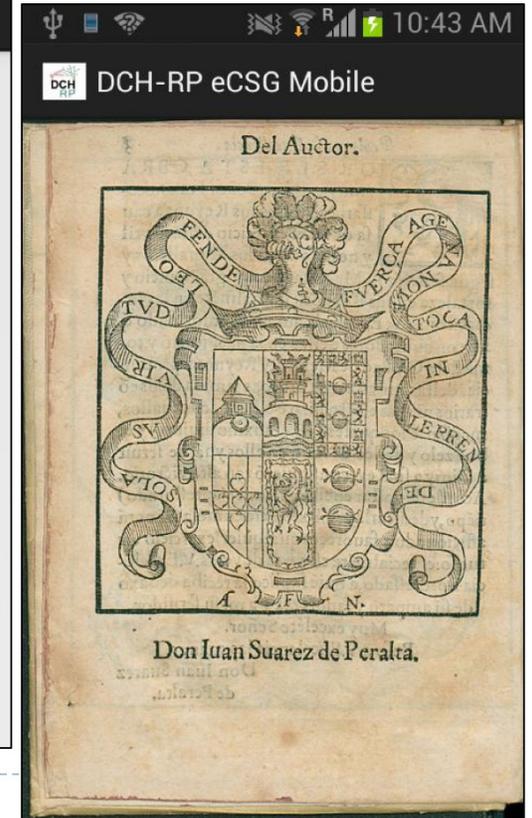
Mini workshop e riunione de

# Attività: Dariah

14



## The DCH-RP e-CSG Mobile



## The DCH-RP e-CSG Mobile



# Attività: Dariah

15

## The DCH-RP e-Culture Science Gateway (<http://ecsg.dch-rp.eu>)

The screenshot displays the DCH-RP e-Culture Science Gateway interface. At the top left, the DCH-RP logo is visible, featuring a stylized tree and the text "DIGITAL CULTURAL HERITAGE ROADMAP FOR PRESERVATION". Below the logo, a "Welcome" message is present. The main interface is divided into several sections:

- SITAR Uploader:** This section is divided into three panels:
  - 1. Collections:** Shows a tree view of "Sitar repositories" with "Origini Informazione" selected.
  - 2. Metadata Browser:** Displays a table of metadata records. The "Records: 0" field is currently empty. The table has columns for "Id\_documento" and "Nome\_file".
  - 3. File Manager:** Shows a "Browse..." button, an "Upload" button, and an "Abort" button. Below these, a list of files is shown, with "1 OL\_1\_DA.pdf" selected. The status bar at the bottom indicates "Selected 1 file(s), 42.1 KB".
- SITAR repository:** This section displays a table of repository records. The table has columns for "Thumb", "nome\_file", "titolo", "zona\_o\_servizio\_ssbar", and "metodo\_georeferenziazion". The records are as follows:

| Thumb | nome_file            | titolo   | zona_o_servizio_ssbar                  | metodo_georeferenziazion |
|-------|----------------------|----------|--|--------------------------|
|       | OI_1066_P7091064.JPG | P7091064 | Municipio XVIII E - Mirella Serlorenzi | base catastale vettorial |
|       | OI_1066_P7091065.JPG | P7091065 | Municipio XVIII E - Mirella Serlorenzi | base catastale vettorial |
|       | OI_1066_P7091072.JPG | P7091072 | Municipio XVIII E - Mirella Serlorenzi | base catastale vettorial |
|       | OI_1066_P7091074.JPG | P7091074 | Municipio XVIII E - Mirella Serlorenzi | base catastale vettorial |
|       | OI_1066_P7091075.JPG | P7091075 | Municipio XVIII E - Mirella Serlorenzi | base catastale vettorial |
|       | OI_1066_P7101081.JPG | P7101081 | Municipio XVIII E - Mirella Serlorenzi | base catastale vettorial |
|       | OI_1066_P7101086.JPG | P7101086 | Municipio XVIII E - Mirella Serlorenzi | base catastale vettorial |

Mini workshop e riunione della C

**SITAR**  
SISTEMA INFORMATIVO TERRITORIALE ARCHEOLOGICO DI ROMA

# Attività: Dariah

16



## Storing and Accessing DARIAH (SADE) contents in EGI

- The overall goal of this mini-project is to create a digital repository of DARIAH contents using  GLibrary
- Datasets for this CC are provided by the Austrian Academic of Science (AAS)
  - Headwords (about 50,000 A-Z) [\[1\]](#)
  - Records (about 40,000 plants; about 70,000 in general) [\[2\]](#)
  - Multimedia with Link to Audio-file (examples; to be improved) [\[3\]](#)
  - Multimedia with Collection (about 3,000; planned to be published within the mini-project) [\[4\]](#)
  - Multimedia connected to Headword (about 3,000; planned to be digitized) [\[5\]](#)
  - Project specific biographies [\[6\]](#)
  - Locations [\[7\]](#)

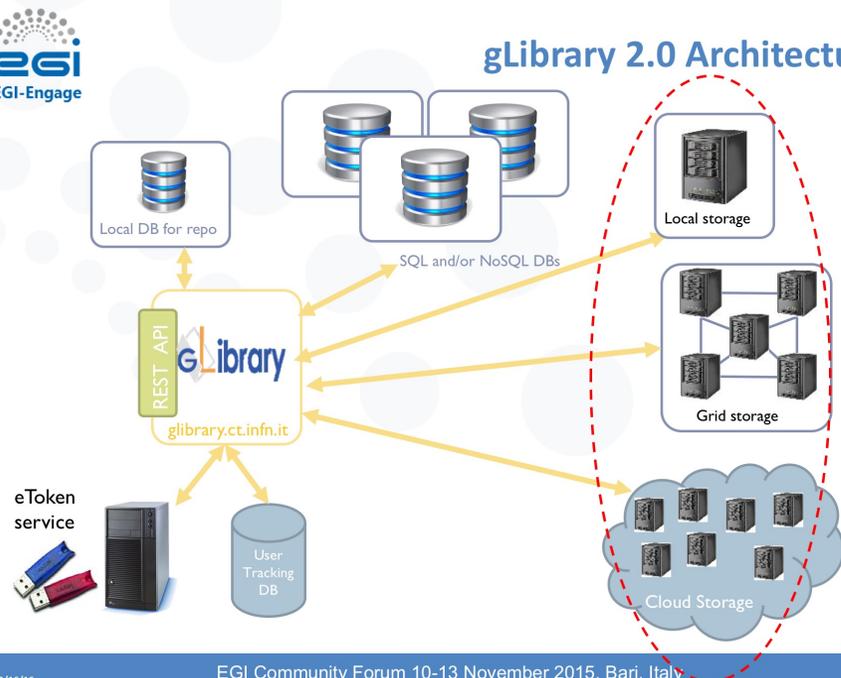
# Attività: Dariah

17



- gLibrary is a **platform** developed by INFN that provides a simple yet powerful system to **organize, search, store** and **retrieve “digital assets”** in distributed repositories built on Grid/Cloud/local storage infrastructures
  - hides the underlying technical details to the users
- “**digital assets**”:
  - digital object + corresponding metadata

gLibrary 2.0 Architecture

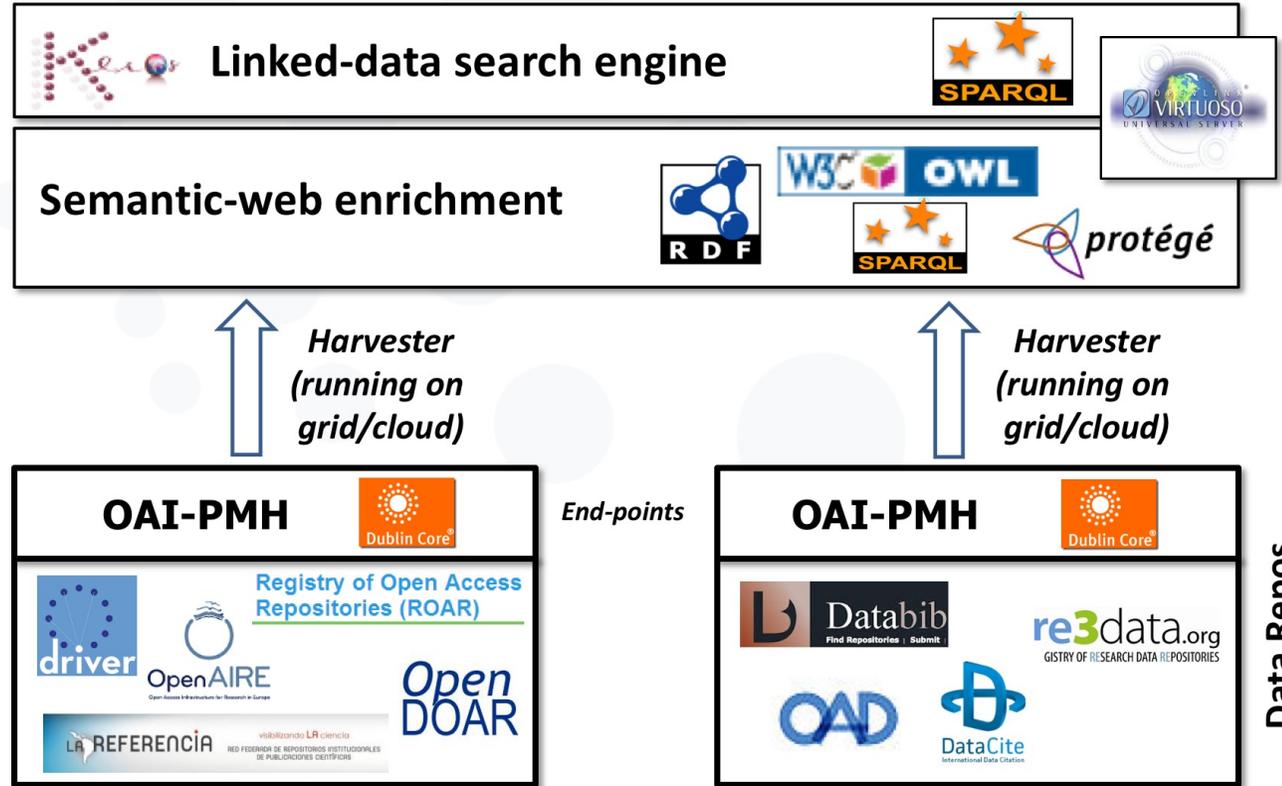


# Attività: Dariah



## The Architecture of the SSE

18



## The Parallel Semantic Search Engine for EGI DARIAH CC

The screenshots show the user interface of the Dariah-CC Parallel Semantic Search Engine. The main search page displays a search bar and a list of search results. A prominent green banner across the results states: **More than 30 million resources, almost 600 million triples**. The search results include a record for "Prvo priopćenje o vrsti Puccinia distincta McAlpine..." with a link to "Portal of scientific journals of Croatia".

At the bottom of the page, there is a footer for the EGI Community Forum, dated 3/16/2015, held in Bari, Italy.

UADKS

16/16

EGI Community Forum 10-13 November 2015, Bari, Italy

Data Repos.

9

# Sinergie con altri progetti

19

- ▶ Molte delle attività mostrate vedono una fase di technology enanchment nei progetti più tecnologici
  - ▶ PRISMA prima, INDIGO-DC adesso
- ▶ e una fase di applicazione agli use case nei progetti più user-driven
  - ▶ BioVeL, Lifewatch-CC, Dariah-CC, Assegni di ricerca e/o convenzioni

# Conclusioni

20

- ▶ È molto importante per l'INFN partecipare alla JRU e/o consorzi nazionali
  - ▶ L'obiettivo è quello di dare in qualche modo sostenibilità ai servizi realizzati nei vari progetti
- ▶ L'INFN è riconosciuta come un partner tecnologico affidabile e in grado di supportare le necessità delle comunità con le più moderne tecnologie
- ▶ I gruppi di ricerca che lavorano in questo contesto hanno già dimostrato che si riesce a stabilire un'ottima sinergia con le attività dei vari progetti in cui partecipano