



# Cloud Computing, Open Science, Data Analytics

Towards new research paradigms for Bavarian dialectology in using advanced research infrastructures

DARIAH – Competence Centre in a nutshell

Eveline Wandl-Vogt, Amelie Dorn, Roberto Barbera & Davor Davidovic



# Cloud Computing, Open Science, Data Analytics

Towards new research paradigms for Bavarian dialectology in using advanced research infrastructures

DARIAH – Competence Centre in a nutshell

Eveline Wandl-Vogt, **Amelie Dorn**, Roberto Barbera & Davor Davidovic



# Outline

1. Introduction  
cultural diversity & European research infrastructures
2. **EGI ENGAGE & the DARIAH Competence Centre (CC)**  
tools, services & resources
3. **Towards joint understanding & collaboration**  
workflows & processes
4. **Outlook**  
organisational design, experimental spaces & future infrastructures



## Introduction

### Cultural and language diversity & biodiversity

- UNESCO – Declaration of Cultural Diversity 2001
- linguistic, cultural & biological diversity: manifestations of diversity of life
- local and indigenous communities
  - complex classification systems reflect understanding of natural environment
  - traditional knowledge maintained in language



## Introduction

### Policies, strategies & infrastructures

UNESCO

cultural diversity

OPEN INNOVATION

curiosity driven collaboration

OPEN SCIENCE

EGI ENGAGE & DARIAH competence  
centre



## Introduction

- Exploitation of advanced research **infrastructures** to discover cultural diversity

→ DARIAH CC & EGI ENGAGE

(Engaging the research communities with open science commons)

→ example

non-standard language resource:

collection of Bavarian dialects in Austria /

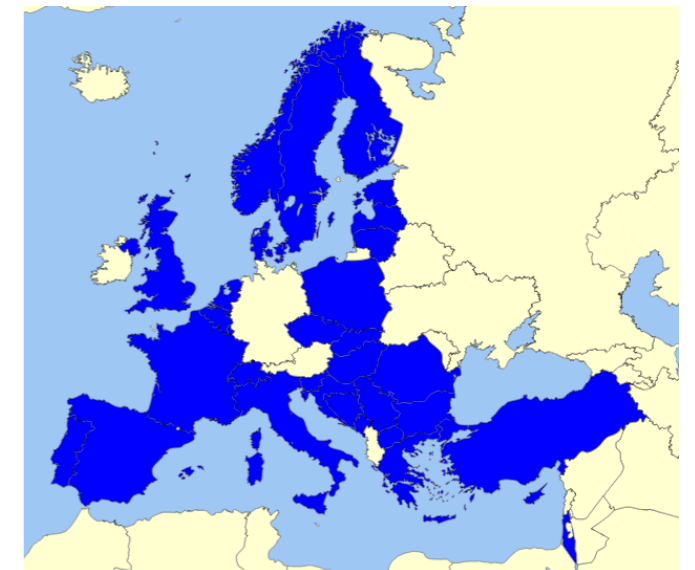
Datenbank der Bairischen Mundarten in Österreich (DBÖ)



## EGI - European Grid Infrastructure



- Federated e-Infrastructure → advanced computing services for research and innovation
- 32 European countries (National grid initiative)
- > 300 data centres & cloud providers
- compute power, storage, applications, clouds etc.



## EGI-Engage Project

- accelerate implementation of Open Science Commons → expanding capabilities of a European backbone of federated services



## DARIAH - Digital Research Infrastructure for Arts and Humanities



- in ~ 17 countries: connecting scholars & research facilities
- supporting digital research → digital tools, sharing data and know-how

### DARIAH Science Gateway

- EGI-service within the EGI-Engage project
- target group: Digital Humanities community







## DARIAH Competence Centre (CC)



- 1 of 8 competence centres established during EU H2020 project EGI-Engage

### contact point

- between EGI & DARIAH communities → advise on infrastructural needs

### customised e-Infrastructure services

- for Arts & Humanities → access to EGI FedCloud resources



## DARIAH Competence Centre (CC) I



demonstrator applications for the Arts & Humanities

- show capabilities/possibilities with the use of EGI

raise awareness

- within scientific Arts & Humanities community on benefits of using e-Infrastructure technologies



# DARIAH CC information via wiki



Page [Discussion](#)





[Read](#) [View source](#) [View history](#)  [Go](#) [Search](#)

## Competence centre DARIAH

EGI-Engage Competence centres: [Main page](#) [ELIXIR](#) [BBMRI](#) [MoBrain](#) [DARIAH](#) [LifeWatch](#) [EISCAT\\_3D](#) [EPOS](#) [Disaster Mitigation |](#) [EGI-Engage Knowledge Commons](#)

### About

The DARIAH Competence Center (CC) is a virtual research group, consisting of computer scientists and the researchers coming from the Arts and Humanities. Its mission is to broadening the usage of the advanced research infrastructures and technologies, such as cloud-oriented services, computational and storage resources in the domain of the Arts and Humanities research. The main activities of the DARIAH CC are:

1. Bringing together relevant stakeholders from the EGI and DARIAH communities who help DARIAH members to better understand their research infrastructure needs, and who operate as a 'knowledge hub' concerning the use of e-Infrastructures in the domain. >> [Read about the CC members](#) 
2. Providing customised e-Infrastructure services for Arts and Humanities, by extending generic EGI e-Infrastructure services according to the unique needs of Arts and Humanities users. >> [Try the e-infrastructure services](#) 
3. Creating demonstrator applications from the Arts and Humanities domain that use e-Infrastructure services. The demonstrators display state-of-the-art capabilities of what's available in digital science with the use of EGI. >> [Learn about and try our demonstrators](#) 
4. Raising awareness within the Arts and Humanities researcher community teams about the benefits of e-Infrastructure technologies. The CC is organising outreach, training and workshop events, and contributes to high-impact events in the field. >> [Check our event calendar](#) 

THE REST OF THE PAGE, INCLUDING THE REFERENCED SERVICES, IS UNDER FINALISATION

### Participants

Ruđer Bošković Institute (RBI), Croatia (Leader institute)	Davor Davidović (CC coordinator), Karolj Skala
Hungarian Academy of Science, Institute for Computer Science and Control (MTA SZTAKI), Hungary	Zoltán Farkas (CC technical coordinator)
Italian National Institute of Nuclear Physics (INFN), Italy	Roberto Barbera, Antonio Calanducci
Gesellschaft für wissenschaftliche Datenverarbeitung mbH (GWDG), Germany	Philipp Wieder, Tibor Kalman
Data Archiving and Networked Services (DANS), Netherlands	Rene van Horik
Austrian Academy of Science (AAS), Austria	Eveline Wandl-Vogt

#### Contents [hide]

- 1 About
- 2 Participants
- 3 Generic services for Digital Arts and Humanities
  - 3.1 Cloud-based compute and storage infrastructure
  - 3.2 DARIAH Science Gateway
  - 3.3 EGI Training Infrastructure
- 4 Services for developers and service providers in Digital Humanities
  - 4.1 gLibrary
  - 4.2 CDSTAR
- 5 Demonstrator applications
  - 5.1 Semantic Search Engine
  - 5.2 Parallel Semantic Search Engine
  - 5.3 DBO@Cloud - The Virtual Dialect Dictionary
  - 5.4 Multi-Source Distributed Real-Time Search and Information Retrieval Application (SIR)
- 6 Events
  - 6.1 Past events
  - 6.2 Future events
- 7 CC project background information
- 8 Tasks
  - 8.1 Task 1: User Support and Training
  - 8.2 Task 2: DARIAH eScience Gateway on EGI
  - 8.3 Task 3: Storing and Accessing DARIAH contents on EGI (SADE)
  - 8.4 Task 4: Multi-Source Distributed Real-Time Search and Information Retrieval (SIR)
  - 8.5 Task 5: Exploitation
- 9 Milestones (M) and Deliverables (D)
- 10 Outputs
  - 10.1 Deliverables
  - 10.2 Virtual organisation
  - 10.3 Science gateways
  - 10.4 Applications
  - 10.5 Survey
- 11 Dissemination activities
  - 11.1 List of publications

- Navigation
- [Main page](#)
  - [Community portal](#)
  - [Current events](#)
  - [Recent changes](#)
  - [Random page](#)
  - [Help](#)

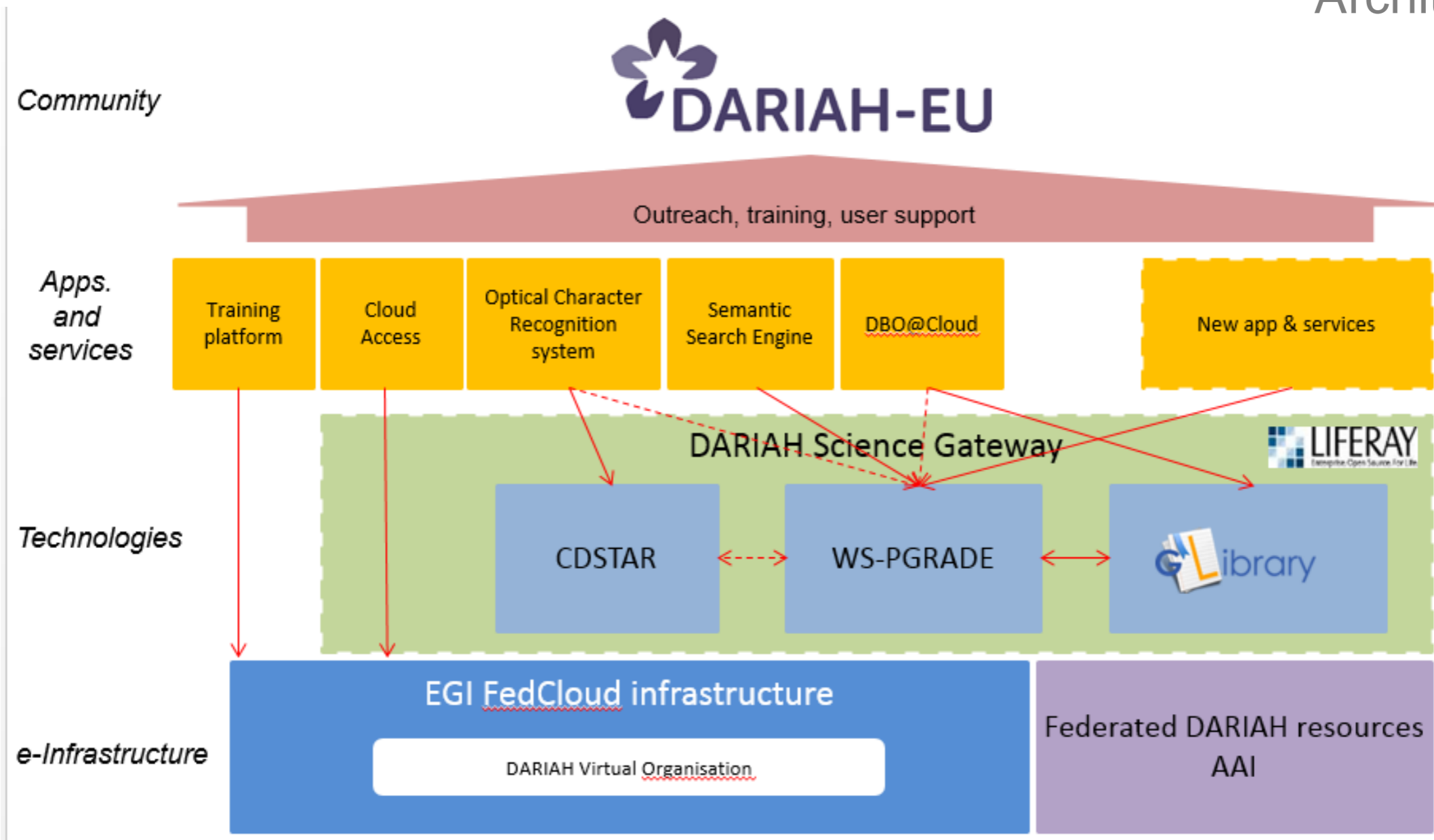
- Toolbox
- [What links here](#)
  - [Related changes](#)
  - [Special pages](#)
  - [Permanent link](#)

- Print/export
- [Create a book](#)
  - [Download as PDF](#)
  - [Printable version](#)



# DARIAH Competence Centre (CC) II

Architecture





## DARIAH Science Gateway

### Applications (open access)

- Semantic Search Engine
  - Simple Semantic Search Engine (SSE)
  - Parallel Semantic Search Engine (PSSE)
- DBO@Cloud

### Services (authentication required)

- Cloud Access
- Workflow Development
- File transfer



# DARIAH CC

## DARIAH science gateway

Getting Started Latest Headlines Imported Latest Headlines Other bookmarks

Sign In

EGI DARIAH CC Gateway

Welcome Applications Cloud Access Workflow Development File Transfer Help

Welcome

## Welcome to the EGI DARIAH Gateway!

This gateway was developed within the [EGI-Engage H2020 project](#) and operated by the [EGI DARIAH Competence Centre](#). This gateway displays state-of-the-art capabilities of what's available in digital science with the use of EGI. The gateway provides various web-based applications and services for the Digital Humanities researchers, institutes and communities.

The gateway includes a [Help](#) page for giving more information on the applications and services it provides.

The gateway includes the following [applications](#):

- **Simple Semantic Search Engine (SSE):**  
Allows users to search in the [e-Infrastructure Knowledge Base](#) ([Open Access Document Repositories](#) and [Data Repositories](#)).  
No authentication is required to access the service.
- **Parallel Semantic Search Engine (PSSE):**  
PSSE is a parallelised version of SSE enabling simultaneously search across the [e-Infrastructure Knowledge Base](#), [Europeana](#), [Cultura Italia](#), [Isidore](#), [OpenAgris](#), [PubMed](#) and [DBpedia](#) platforms.  
No authentication is required to access the service.
- **DBO@Cloud:**  
A Cloud-based repository presenting 100+ years old collection of Bavarian dialects.  
No authentication is required to access the service.

The gateway includes the following [services](#):

- **Cloud Access:**  
Single-job applications and parameter-sweep applications can be run on the DARIAH VO clouds without porting efforts.  
[Authentication](#) is required to access the service.
- **Workflow Development:**  
Complex workflow applications can be developed and run on all the resources of the DARIAH VO.  
[Authentication](#) is required to access the service.
- **File transfer:**  
Enables transferring data from, to and between storage services providing HTTP, HTTPS, SFTP, GSIFTP, SRM, iRODS and S3 protocols.  
[Authentication](#) is required to access the service.



## (Simple) Semantic Search Engine I


- Division of Catania of the National Institute of Nuclear Physics (INFN)
- Semantically enriches the metadata of the resources included in the Knowledge Base
- Queries in more than 110 languages possible
- Results connected to Google Scholar and Altmetric if available
- Uses LODLive API to explore the Linked Data graph
- Extension: Parallel Semantic Search Engine (PSSE)  
<https://dariah-gateway.lpds.sztaki.hu/web/guest/psse>

ÖAW



# Semantic Search Engine II

SEMANTISEARCH-PORTLET

 Examples  Search

Enter a keyword

Results

Records found. Displaying 1 to 20

Title

Possible implementation of e-Infrastructures for Cultural Heritage (Citations) (Altmetrics) (Linked Data)

Author

*Barbera R; Barbera R; Maggi M*

Description

Presentation given at the international conference "Research Infrastructure and e-Infrastructures for Digital Cultural Heritage" - Rome, 13-14 November 2014

Repository

[Open Access Repository](#)

[More Info](#)

Title

Book Review of *Archaeological Theory and the Politics of Cultural Heritage* by Laurajane Smith (Citations) (Linked Data)

Author

*Nicholaus George P.*

Description

The passage of the Native American Graves Protection and Repatriation Act (NAGPRA) in 1991 significantly changed the way archaeology would be done in the United States. This act was presaged by growing complaints and resentment directed at the scientific community by Native Americans over the treatm...

Repository

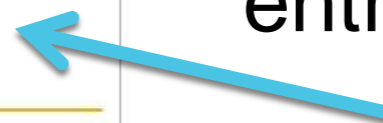
[DigitalCommons@University of Nebraska](#)

[More Info](#)

Explore repositories

- keyword(s) for searching

- result list of data entries





ÖAW



# Semantic Search Engine III

SEMANTISEARCH-PORTLET

## Possible implementation of e-Infrastructures for Cultural Heritage

### General Information

**Authors:** B a r b e r a R; Barbera R; Maggi M

**Description :** Presentation given at the international conference Research Infrastructure and e-Infrastructures for Digital Cultural Heritage - Rome, 13-14 November 2014

**Identifier :** <http://www.openaccessrepository.it/record/9817/files/PRESENTATIONINFN-2014-025.pptx>

**Identifier :** <http://www.openaccessrepository.it/record/9817>

**Identifier :** info:doi:10.15161/oar.it/1415951975.9

**Language :** eng

**Rights :** cc-by

**Audience :** Researchers

### Linked Data

### Date Information <sup>↗</sup>

**DateStamp:** 2015-09-24T08:31:58Z

**Date :** 2014-11-14T08:03:49Z

### Dataset Information <sup>↗</sup>

### Repository Information <sup>↗</sup>

### Information from Google Scholar <sup>↗</sup>

### Altmetrics <sup>↗</sup>

Detailed information of the selected record from the result list



# DBO@Cloud

- data originally collected for a project at the Austrian Academy of Sciences (Dictionatry/Database of Bavarian Dialects in Austria)
- > 20 000 records with metadata can be browsed  
<https://dariah-gateway.lpds.sztaki.hu/web/guest/dbo-portlet>
- Filter options and customizing of the display
- Images from the original questionnaires available

The screenshot shows the DARIAH Browser interface. On the left, there is a table with columns: Titel, Autor, Kurzzeit, Langzeit, and Erscheinungsjahr. The table lists various materials with their respective authors and years. On the right, there is a 'Preview Window' showing a handwritten document titled '9. Ergänzungsfragebogen' (Supplementary Questionnaire) from 1935. The document contains a list of questions in German, such as 'Woher sind Sie?', 'Wie alt sind Sie?', and 'Woher kommen Sie?'. The answers are handwritten in cursive.



## Workflows & Processes

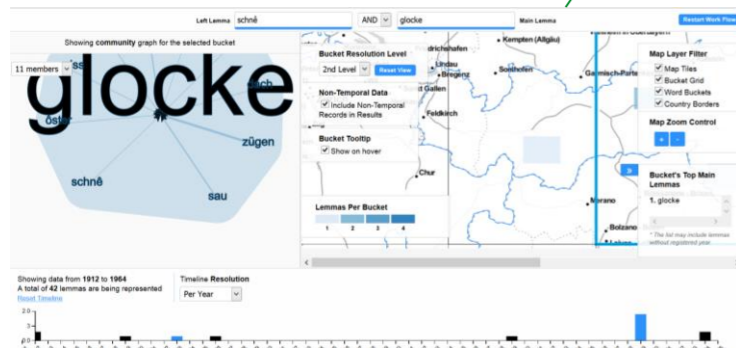
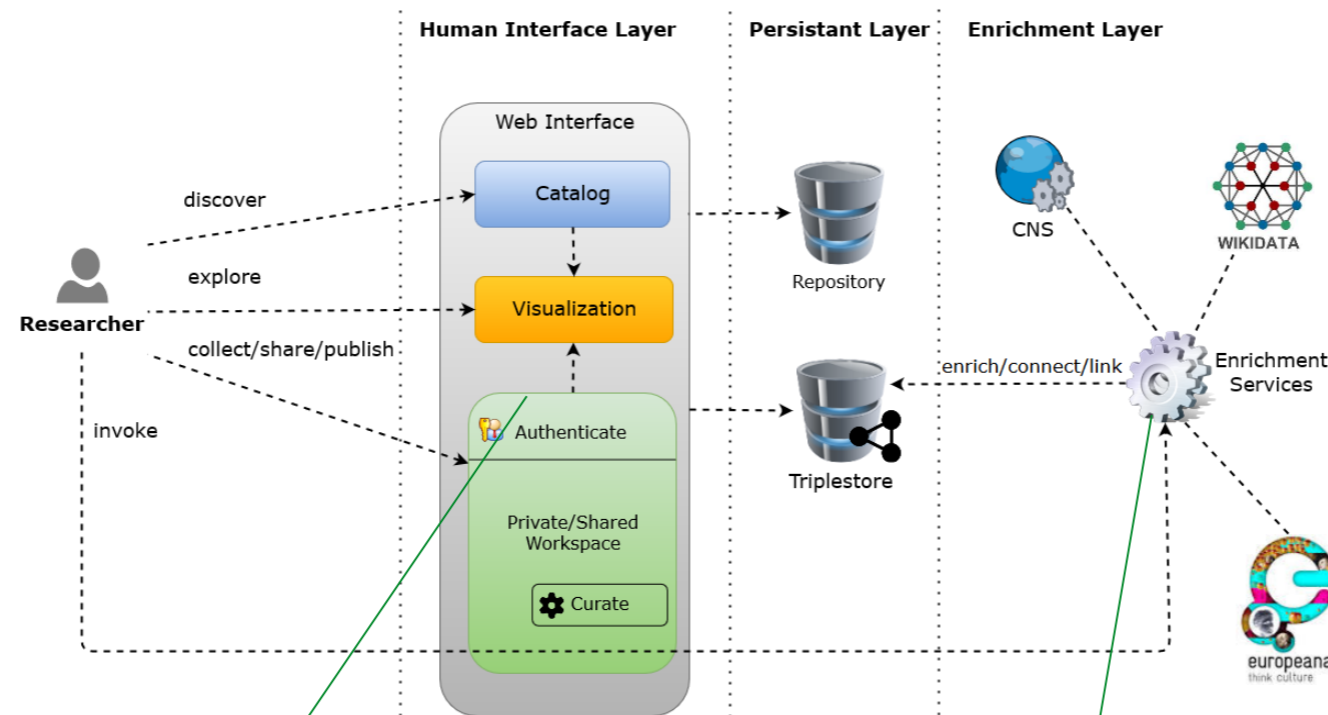
### Towards joint understanding & collaboration

- Multidisciplinarity  
communication & knowlege exchange processes
- Multilinguality /Multiculturality  
collaboration in a European team with different cultures
- Modularity



# DARIAH CC – added value for cultural lexicography

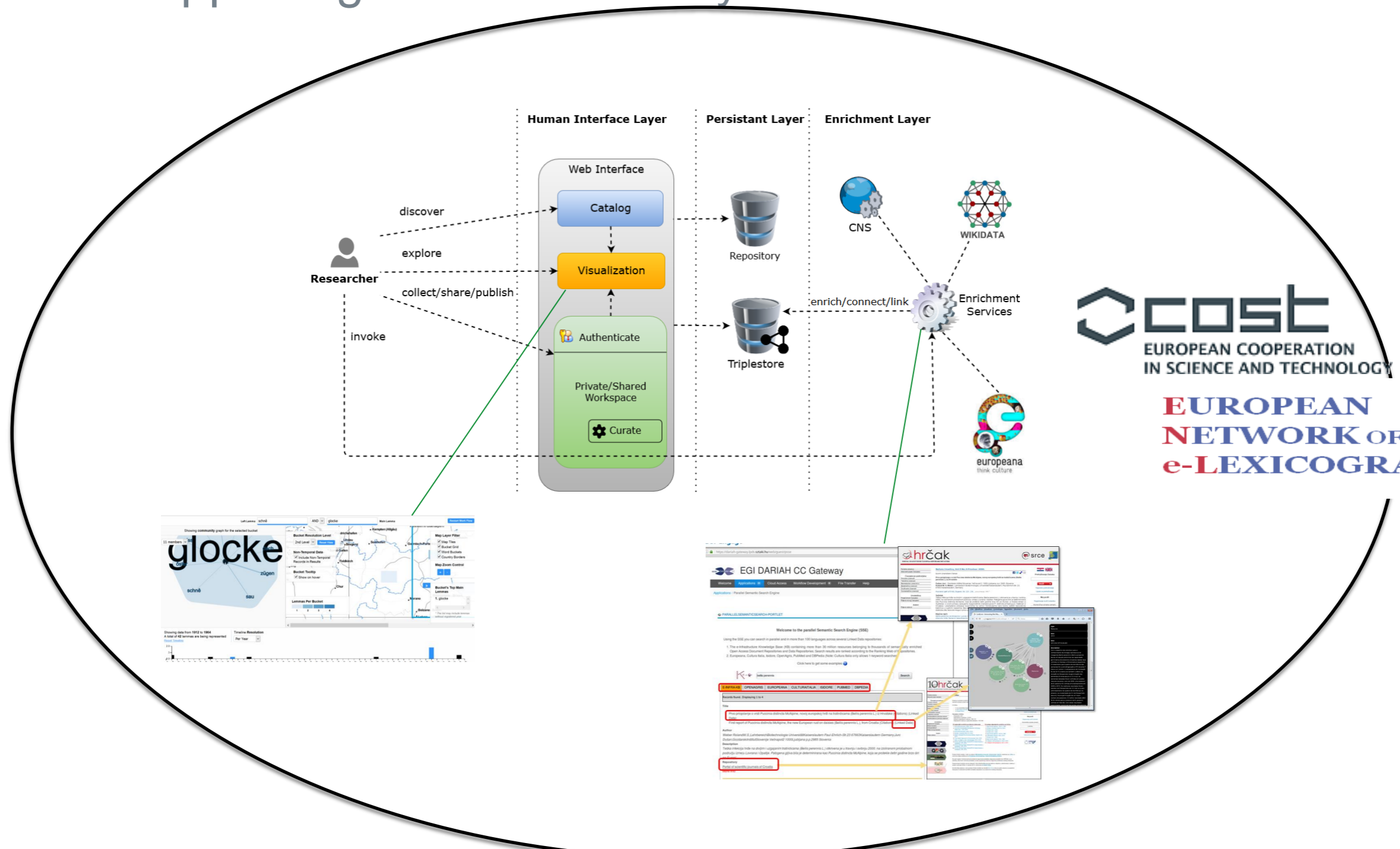
## supporting cultural diversity documentation & research



This block contains three screenshots:

- EGE DARIAH CC Gateway:** A navigation menu with options like "Applications", "Cloud Access", "Workflow Development", and "File Transfer".
- Parallel Semantic Search Engine (SSE):** A search interface with a search bar and a list of search results.
- Search Result:** A detailed view of a search result for a scientific paper titled "Prvo priopćenje o virusu Puccinia disticta McAlpine, novog europskog vrsta na tratinčama (Bellis perennis L.) iz Hrvatske". The result includes the author's name, a description of the paper, and the repository information.

supporting cultural diversity documentation and research



## Learnings from a case study approach

- re-modeling of a long term humanities project is possible  
yet challenging
- goes along with micro-prototyping approach
- fits with Open Innovation in Science methods

digital transformation  
collaborative organisations

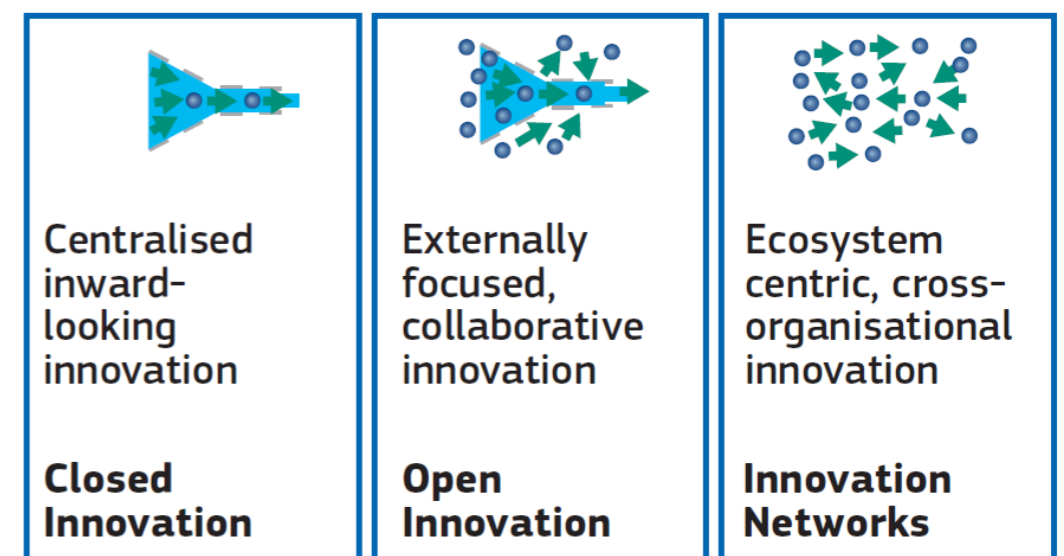
# Outlook

## Organisational design, experimental spaces & innovation

- virtual, open collaboration
- common goals and passion
- designing workflows, opening up

innovation networks  
ecosystem centric

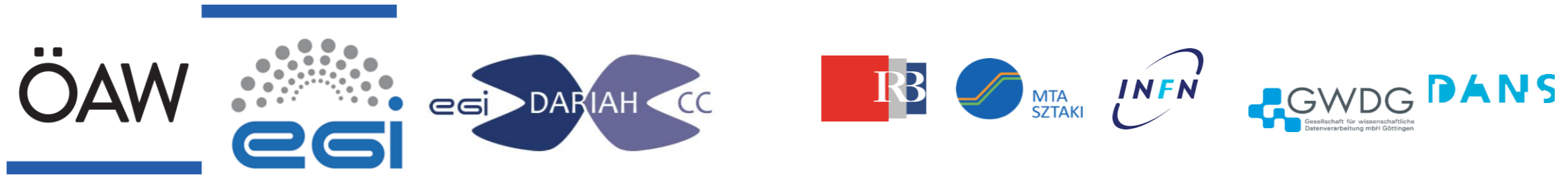
Figure 6: The evolution of innovation





Thank you!





## Bibliography

- DARIAH Competence Center EGIWiki  
[https://wiki.egi.eu/wiki/Competence\\_centre\\_DARIAH](https://wiki.egi.eu/wiki/Competence_centre_DARIAH) (accessed 10.08.2017)
- EGI ENGAGE Mainpage  
[https://wiki.egi.eu/wiki/EGI-Engage:Main\\_Page](https://wiki.egi.eu/wiki/EGI-Engage:Main_Page) (accessed 10.08.2017)
- EGI DARIAH CC Gateway  
<https://dariah-gateway.lpds.sztaki.hu/> (accessed 10.08.2017)
- Wandl-Vogt, E., Barbera, R., La Rocca, G., Calanducci, A., Carrubba, C., Inserra, G., Kalman, T., Sipos, G., Farkas, Z. & Davidović, D. (2016) Furthering the Exploration of Language Diversity and Pan-European Culture: The DARIAH-CC Science Gateway for Lexicographers. Proceedings of EURALEX2016, Tbilisi, Georgia. pp.284-290