

Hands-on: Znanstveni portal za DARIAH zajednicu

Davor Davidović, Eva Cetinić

Ruđer Bošković Institute
DARIAH Competence Centre, EGI-Engage
INDIGO-DataCloud



Prezentacija dostupna na:

<http://bit.ly/2t0socz>

DARIAH Znanstveni portal

- Rezultat DARIAH Centra kompetencije (EGI-Engage projekt)
- Omogućiti znanstvenicima pristup Cloud infrastrukturi (EGI FedCloud)
 - jednostavno i bez potrebe za tehničkim predznanjem
- Trenutno nudimo nekoliko alata i servisa
 - Većinu ćemo proći kroz ovu vježbu!

DARIAH CC servisi



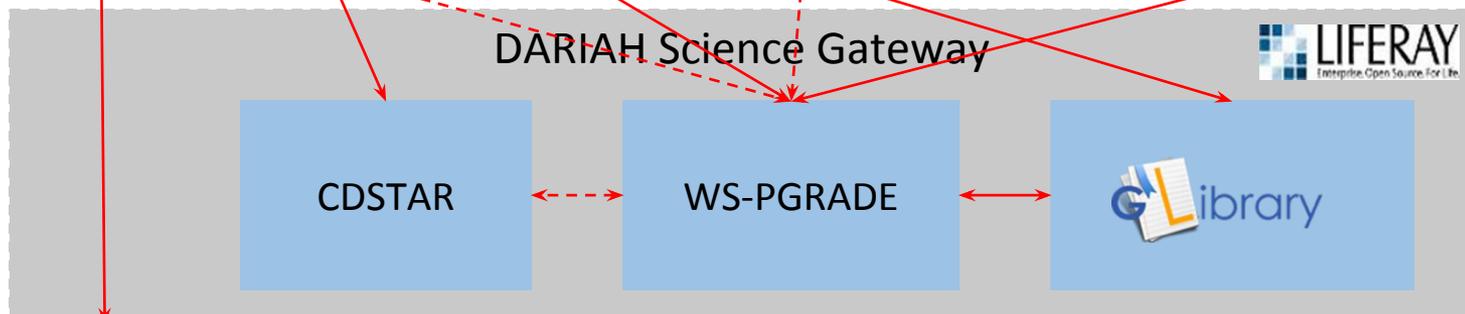
Community

Outreach, training, user support

Apps.
and
services



Technologies



e-Infrastructure



DARIAH Znanstveni portal

Vježba 1: Upoznavanje s portalom

Upoznavanje s portalom

Cilj vježbe:

- Naučiti kako pristupiti i registrirati se/prijaviti na portal te upoznati osnovnu funkcionalnost portala

Što ćemo napraviti:

- Prijavit se na portal koristeći naš EduHr korisnički račun
- Upoznati se s izgledom i osnovnim funkcijama portal

DARIAH Znanstveni portal

Link: <https://dariah-gateway.lpds.sztaki.hu/>

← → ↻ Sigurno | <https://dariah-gateway.lpds.sztaki.hu>

EGI DARIAH CC Gateway Sign In

Welcome Applications Cloud Access Workflow Development File Transfer Help Forum

Welcome

Web Content Display

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 gateway service and then robot certificate provides access to the cloud.
- **Workflow Development:**
Complex workflow applications can be developed and run on all the resources of the DARIAH VO.
[Authentication](#) is required to access the gateway service and then robot certificate provides access to the cloud.
- **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 gateway service and then additional authentication is needed to access the selected storage service.

Web Content Display

DARIAH Science Gateway is an EGI service provided by DARIAH Competence Centre, co-funded by EGI-Engage <http://go.egi.eu/eng>

Prijava na portal

Prijava na portal



- Welcome
- Applications
- Cloud Access
- Workflow Development
- Information
- File Transfer
- Help
- Admin
- Forum

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 gateway service and then robot certificate provides access to the cloud.
- **Workflow Development:**
Complex workflow applications can be developed and run on all the resources of the DARIAH VO.
Authentication is required to access the gateway service and then robot certificate provides access to the cloud.
- **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 gateway service and then additional authentication is needed to access the selected storage service.



DARIAH Science Gateway is an EGI service provided by DARIAH Competence Centre, co-funded by EGI-Engage <http://go.egi.eu/eng>

Powered By [Liferay](#)

Odaberite svog ID providera

DFN-AAI + eduGAIN

eduGAIN **DFN**
Deutsches
Forschungsnetz

[About eduGAIN](#) | [Help](#)

Select your organisation

In order to access the service **darjah-gateway.lpd.sztaki.hu** please select or search the organisation you are affiliated with.

AAI@EduHr Single Sign-On Service

Remember selection for this web browser session.
 Remember selection permanently and bypass this step from now on.

[Impressum](#) Software provided by [SWITCH](#)

Odaberite
AAI@EduHr
i
"Select"

Korisničko ime i lozinka

AA
@EduHr

Korisnička oznaka
ddavid@irb.hr

Zaporka
.....

[Prijavi se](#) [Pomoć](#)

Autentikacijska i autorizacijska infrastruktura znanosti i visokog obrazovanja u Republici Hrvatskoj

Korisničko ime

Lozinka

Početna stranica – nakon spajanja



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 gateway service and then robot certificate provides access to the cloud.
- **Workflow Development:**
Complex workflow applications can be developed and run on all the resources of the DARIAH VO.
[Authentication](#) is required to access the gateway service and then robot certificate provides access to the cloud.
- **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 gateway service and then additional authentication is needed to access the selected storage service.



DARIAH Science Gateway is an EGI service provided by DARIAH Competence Centre, co-funded by EGI-Engage <http://go.egi.eu/eng>

DARIAH Znanstveni portal

Vježba 2: Semantička pretraga baza podataka

Semantic Search Engine (SSE)

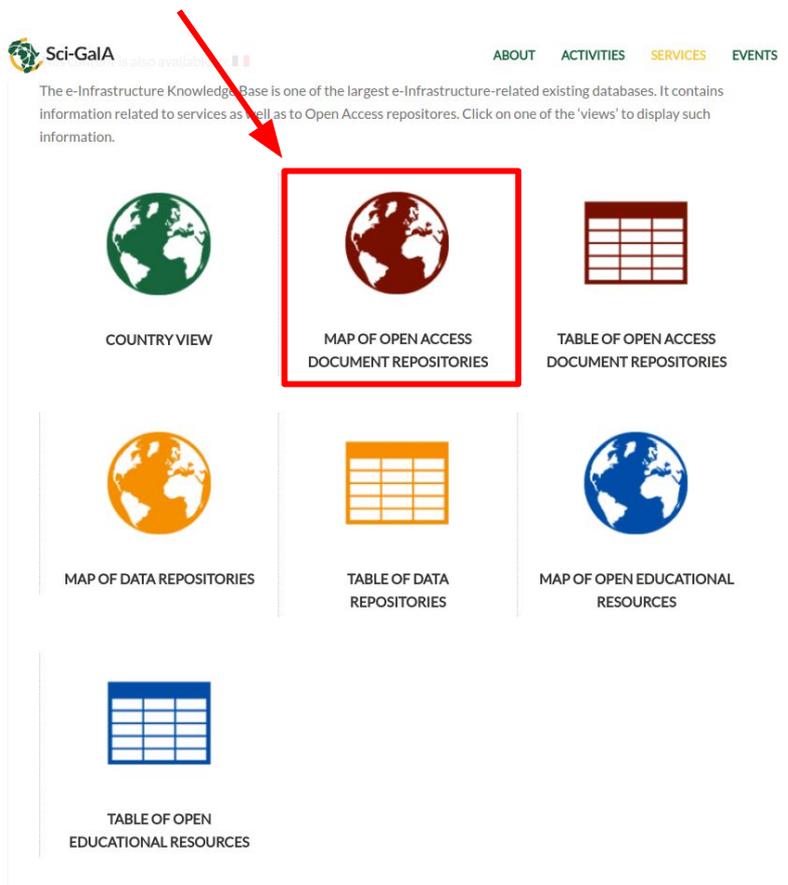
- Razvijen od Nacionalnog instituta za nuklearnu fiziku, Catania (INFN), Italija
- pretražuje e-Infrastructure Knowledge Base s više od 30 milijuna resursa
<http://www.sci-gaia.eu/knowledge-base/>
- mogućnost pretrage u više od 110 jezika
- rezultati pretrage se rangiraju prema [Ranking Web of Repositories](#)

Pretraga

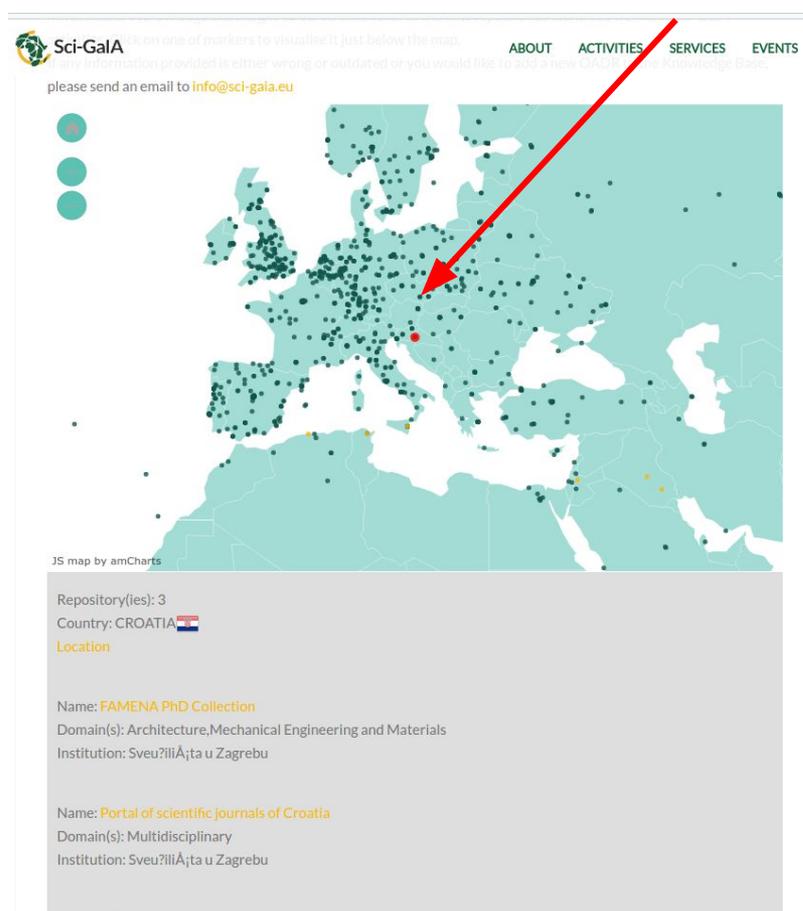
1. Otvorite e-Infrastructure Knowledge Base

2.

3. Hrvatska



The screenshot shows the Sci-GaIA homepage. At the top, there is a navigation menu with 'ABOUT', 'ACTIVITIES', 'SERVICES', and 'EVENTS'. Below the menu, a red arrow points to the 'MAP OF OPEN ACCESS DOCUMENT REPOSITORIES' option, which is highlighted with a red box. Other options include 'COUNTRY VIEW', 'TABLE OF OPEN ACCESS DOCUMENT REPOSITORIES', 'MAP OF DATA REPOSITORIES', 'TABLE OF DATA REPOSITORIES', 'MAP OF OPEN EDUCATIONAL RESOURCES', and 'TABLE OF OPEN EDUCATIONAL RESOURCES'.



The screenshot shows the Sci-GaIA map interface. At the top, there is a navigation menu with 'ABOUT', 'ACTIVITIES', 'SERVICES', and 'EVENTS'. Below the menu, there is a text prompt: 'please send an email to info@sci-gala.eu'. The main part of the interface is a map of Europe with numerous green dots representing repositories. A red arrow points to a red dot in Croatia. Below the map, there is a panel with the following information:

Repository(ies): 3
Country: CROATIA 
Location

Name: **FAMENA PhD Collection**
Domain(s): Architecture, Mechanical Engineering and Materials
Institution: Sveučilište u Zagrebu

Name: **Portal of scientific journals of Croatia**
Domain(s): Multidisciplinary
Institution: Sveučilište u Zagrebu

Pretraga prema autoru

1. Upišite “author: Michelangelo”

SEMANTISEARCH-PORTLET

The screenshot shows the SEMANTISEARCH-PORTLET search interface. At the top, there is a search bar with the text "author: michelangelo" and a "Search" button. Below the search bar, there is a "Results" section with a yellow header. The results are displayed in a list format, showing the first two results. Each result includes a title, author information, a description, and repository information. The first result is titled "Transductive relational classification in the co-training paradigm (Citations) (Altmetrics) (Linked Data)" and lists authors: Guo Hongyu; Paquet Eric; Viktor Hernal; Ceci Michelangelo; Appice Annalisa; Malerba Donato. The second result is titled "Network regression with predictive clustering trees (Citations) (Linked Data)" and lists authors: Ceci Michelangelo; Appice Annalisa; Dzeroski Saso; Stojanova Daniela. Both results include a "More Info" link.

Records found. Displaying 1 to 20

Title
Transductive relational classification in the co-training paradigm (Citations) (Altmetrics) (Linked Data)

Author
Guo Hongyu; Paquet Eric; Viktor Hernal.; Ceci Michelangelo; Appice Annalisa; Malerba Donato

Description
peer reviewed: yes

Repository
NRC Publications Archive
[More Info](#)

Title
Network regression with predictive clustering trees (Citations) (Linked Data)

Author
Ceci Michelangelo; Appice Annalisa; Dzeroski Saso; Stojanova Daniela

Description
Network data describe entities represented by nodes, which may be connected with (related to) each other by edges. Many network datasets are characterized by a form of autocorrelation, where the value of a variable at a given node depends on the values of variables at the nodes it is connected with....

Repository
[Open Access Infrastructure for European Research](#)
[More Info](#)

Probajte s drugim pojmovima, npr.

- “linked data”
- “subject: musicology”
- “subject: second world war”
- “author: vaše prezime”

Parallel Semantic Search Engine (PSSE)

The image displays the Parallel Semantic Search Engine (PSSE) interface. The main search results page shows a search for "bellis perennis" with records found from various repositories. A red box highlights the first result: "Prvo priopćenje o vrsti Puccinia distincta McAlpine, novoj europskoj hrđi na tratinčicama (Bellis perennis L.) iz Hrvatske (Citations) (Linked Data)". Another red box highlights the repository information: "Repository Portal of scientific journals of Croatia".

The detailed view of the search result from the "10hrčak" portal includes the following information:

- Title:** Prvo priopćenje o vrsti Puccinia distincta McAlpine, novoj europskoj hrđi na tratinčicama (Bellis perennis L.) iz Hrvatske (Citations) (Linked Data)
- Author:** Weber Roland W. S., Lehrbereich Biotechnologie Universität Kaiserslautern Paul-Ehrlich-Str. 23 67663 Kaiserslautern Germany; Juro Dušan; Gozdarski inštitut Slovenije Večnapt2 1000Ljubljana p.p.2985 Slovenia
- Description:** Teška infekcija hrđe na divljim i uzgajanim tratinčicama (Bellis perennis L.) otkrivena je u travnju i svibnju 2000. na izoliranom probalnom području između Lovrana i Opatije. Patogena gljiva bila je determinirana kao Puccinia distincta McAlpine, koja se protekle četiri godine brzo širi po Europi.
- Repository:** Portal of scientific journals of Croatia

The interface also shows a navigation menu with options like "Welcome", "Applications", "Cloud Access", "Workflow Development", "File Transfer", and "Help". The search results are displayed in a table with columns for "Title", "Author", and "Description". The detailed view includes a "Records found. Displaying 1 to 4" section and a "Repository" section with a "More Info" link.

DARIAH Znanstveni portal

Vježba 3: Jednostavna analiza podataka

Analiza e-knjiga iz kolekcije Gutenberg

Cilj vježbe:

- Pokazati mogućnost korištenja Cloud infrastrukture za računalno zahtjevnije probleme

Što ćemo napraviti:

- Jednostavnu analizu teksta svih e-knjige određenog autora objavljenih u kolekciji Gutenberg - <https://www.gutenberg.org>

Koraci analize

1. Spojiti se na bazu Gutenberg
2. Za danog autora (ime i prezime) skinuti sve objavljene e-knjige
3. “Očistiti” tekst od suvišnih dijelova (npr. zaglavlja)
4. Analizirati skinute tekstove (python programski jezik)
 - Frekvencija riječi
 - Broj tokena
 - Lingvistička raznolikost
 - Naći riječi duže do 7 znakova koje se pojavljuju barem 10 puta

Što ćemo koristiti za provedbu analize?

- Simple Cloud Access servis
 - Web-servis za brzo i jednostavno pokretanje programa u računalnom oblaku
- Program za analizu teksta
 - <https://bitbucket.org/davordavidovic/textanalysis/overview>
 - Python programski jezik



Krenimo s analizom!

1. Logirati se na portal (vježba 1)
2. U gornjem izborniku odabrati “Cloud Access”
3. Odabrati “Gutenberg demo” ili na link:
<https://bitbucket.org/davordavidovic/textanalysis/overview>
4. Dalje nastavljamo prema uputstvima na web-u

DARIAH Znanstveni portal

Vježba 4: Pretraga

Pretraga DBÖ repozitorija u računalnom oblaku

Cilj vježbe:

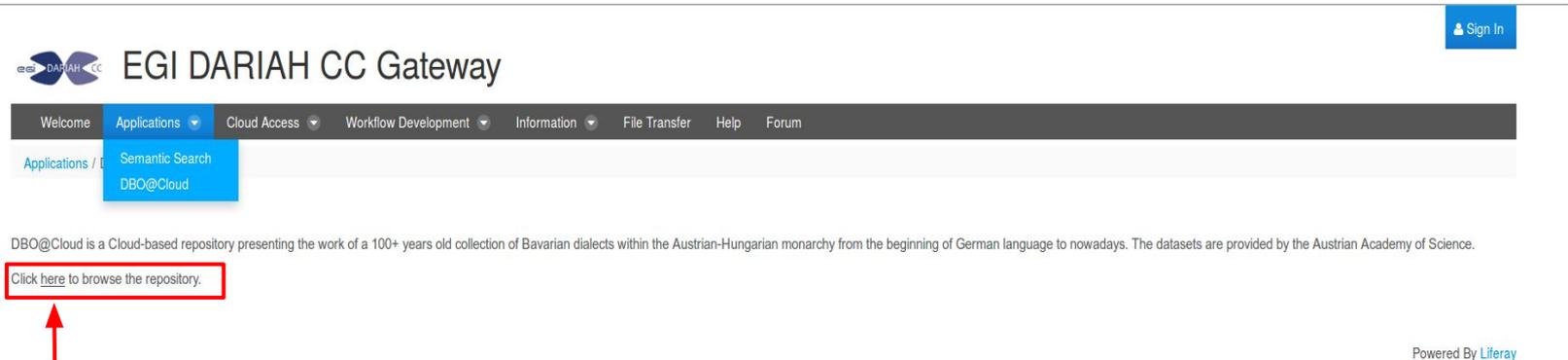
Pokazati na konkretnom primjeru kao se postojeće kolekcije podataka mogu pohraniti i pretražiti koristeći računalni oblak

Što će napraviti:

Pretražit ćemo 100 godina staru bazu podataka Bavarskih dijalekata iz doba Austro-ugarske monarhije (podaci su od Austrijske akademije znanosti)

DBO@Cloud

1. Nije potrebna registracija (ako jeste, nema veze :)
2. U glavnom izborniku: Applications -> DBO@Cloud



The screenshot shows the EGI DARIAH CC Gateway website. The header includes the logo and the text "EGI DARIAH CC Gateway". A navigation bar contains links for "Welcome", "Applications", "Cloud Access", "Workflow Development", "Information", "File Transfer", "Help", and "Forum". The "Applications" menu is open, showing "Semantic Search" and "DBO@Cloud". Below the navigation bar, there is a paragraph of text: "DBO@Cloud is a Cloud-based repository presenting the work of a 100+ years old collection of Bavarian dialects within the Austrian-Hungarian monarchy from the beginning of German language to nowadays. The datasets are provided by the Austrian Academy of Science." A red box highlights the text "Click here to browse the repository." with a red arrow pointing to it from below. The text "Powered By Liferay" is visible in the bottom right corner.

KLIK!

DBO@Cloud

The screenshot shows the EGI DARIAH CC Gateway interface. The main navigation bar includes 'Welcome', 'Applications', 'Cloud Access', 'Workflow Development', 'File Transfer', and 'Help'. The 'Applications' menu is active, showing 'DARIAH Browser' and 'DBO Portlet'. The 'DARIAH Browser' section displays a list of publications with columns for 'Titel' and 'Autor'. The selected entry is 'Schriftsprache und Dialekte im Deutschen nach Zeugnissen alter und neuer Zeit' by Socin, Adolf. A detailed metadata view on the right shows the following information:

Id:	32
Titel:	Schriftsprache und Dialekte im Deutschen nach Zeugnissen alter und neuer Zeit
Autor:	Socin, Adolf
Untertitel:	Beiträge zur Geschichte der deutschen Sprache
Nachdruck:	Reprografischer Nachdruck der Ausgabe Heilbronn 1888
Erscheinungsort:	Hildesheim, New York
Erscheinungsjahr:	1970
Erstauflage:	1888
Verlag:	Georg Olms Verlag
Seitenstruktur:	XII+544
Kurzzeitart:	-

> 20,000 zapisa!

Titel	Autor	Kurztitel	Langtitel
Die Woszbung Almbins von Kärnten	Svetina, Robert		
@cs Das Jagdbuch des Martin Strasser von Kollnitz		Strasser (1624/1997)	Das Jagdbuch des Martin Strasser von Koll...
Metalinguistici al femminile nella ricerca linguistica a Sappada (per il dizionario cappa)	Hornung, Maria		
Das besondere Museum	Hornung, Maria		
Ladinia et Romania			
Personennamen und -identität			
Ir sult sprechen willekommen			
Lexikalische Studien zum Dialekt im Wiener Volksstück vor Nestroy	Reutner, Richard		
Die Mundartdichtung in Salzburg	Rettenbacher, August/Rettenbacher, Barbara		
Die Überlieferung der Ranshofener Traditionen	Schmidt, Rudolf Wolfgang		
Beiträge zur älteren Besitz- und Rechtsgeschichte steirischer Klöster	Pirchegger, Hans		
Geschichte des Bezirks Gröbming	Pirchegger, Hans		
Burgengründer	Neweklowsky, Walter		
Die Linzer Fernstrassen	Pfeffer, Franz		
Dizionario toponomastico del Friuli-Venezia Giulia	Frau, Giovanni		
Die Kompositen in gallischen Personennamen	Schmidt, Karl Horst		
Gaulish Personal Names	Evans, D. Ellis		
Das ehemalige Cisterciensinnen-Kloster zu St. Bernhard bei Horn	Endl, Friedrich		
Salzburg und Hamburg Erzbistumsgründung und Missionspolitik in karolingischer Zeit	Wavra, Brigitte		
Karantainen und der Alpen-Adria-Raum im Frühmittelalter			
Quellenverzeichnis zu den mittelhochdeutschen Wörterbüchern	Nellmann, Eberhard		
---	Anreiter, Peter		
Necrologium Curiense			
Die Gründung des Klosters Maria-Zell im Wiener Wald und die Besitzgeschichte sein...	Lechner, Karl		
Schriftsprache und Dialekte im Deutschen nach Zeugnissen alter und neuer Zeit	Socin, Adolf		
Althochdeutsches Glossenwörterbuch	Wells, John C.		
Régi Magyar Családnevek Szótára	Miklós, Kázmer		
Krankheit und Heilen	Grabner, Elfriede		
Zur Textgeschichte der Vita Ruperti	Beumann, Helmut		
Wie alt ist Klagenfurt?	Webering, Evelyne		
Namenkundliches aus dem Bereich der Landeshauptstadt Klagenfurt	Pohl, Heinz Dieter		
Lexikon deutschsprachiger Mundartautoren	Sowinski, Bernhard	SOWINSKI: (1997)	
Dizionario di Toponomastica	Gasca Queirazza, Giuliano/Marcatto, Carla/Pellegrini, Giov...		
Österreichische Redewendungen und Redensarten	Malygin, Viktor T.		

Detalji zapisa

Id:	2
Typ id:	1
Veroeffentlich:	1
Titel:	@cs Das Jagdbuch des Martin Strasser von Kollnitz
Band:	3. Bd.
Reihentitel:	Das Kärntner Landesarchiv. Studien zur geschichtlichen Landeskunde Kärntens
Auflage:	2., unveränd. Aufl.
Erscheinungsort:	Klagenfurt
Erscheinungsjahr:	1997
Verlag:	Verlag des Kärntner Landesarchivs
Isbn:	3-900531-35-8
Seitenstruktur:	419
Herausgeber:	Lindner, Kurt
Kurztitel:	Strasser (1624/1997)
Kurztitel Sort:	Strasser (1624/1997)
Langtitel:	Das Jagdbuch des Martin Strasser von Kollnitz. Hg. von Lindner, Kurt. 2., unveränd. Aufl. Klagenfurt 1997 (= Das Kärntner Landesarchiv. Studien zur geschichtlichen Landeskunde Kärntens 3)
Freigabe:	1
Checked:	1
Online:	1
Anmerkung:	ANM_BIZ: aus dem Jahre 1624 (Ain Puech von allerlai Jägerrei und Waidmanschaften). ANM_EXZ: Exz. 0 == 419 Seiten == Lf: Gewerkensohn aus Gastein Sa., oberster sa. Jägermeister, 1600 nach Kollnitz in Gm. St.Paul Kä. auf Erzbisch.Besitz ANM_S: Kü.2002/01: Eintr.ok ## Sig. ergänzen, Geb.Zuordng.??, literarisch? M.G.: SL/A: 2./4. Anmerkung: weitere Bände des "Kärntner Landesarchivs" unter N-8189 und A-7135

Category	File Type	File Name	Container Path	File Size	Upload...
No items to show.					

DARIAH Znanstveni portal

Vježba 5 (demo):

Cloud-based repository platform



INDIGO - DataCloud

Cilj:

Pokazati prednosti korištenja cloud i virtualizacijskih tehnologija za brzu dobivanje vlastitog repozitorija

Što ćemo napraviti*:

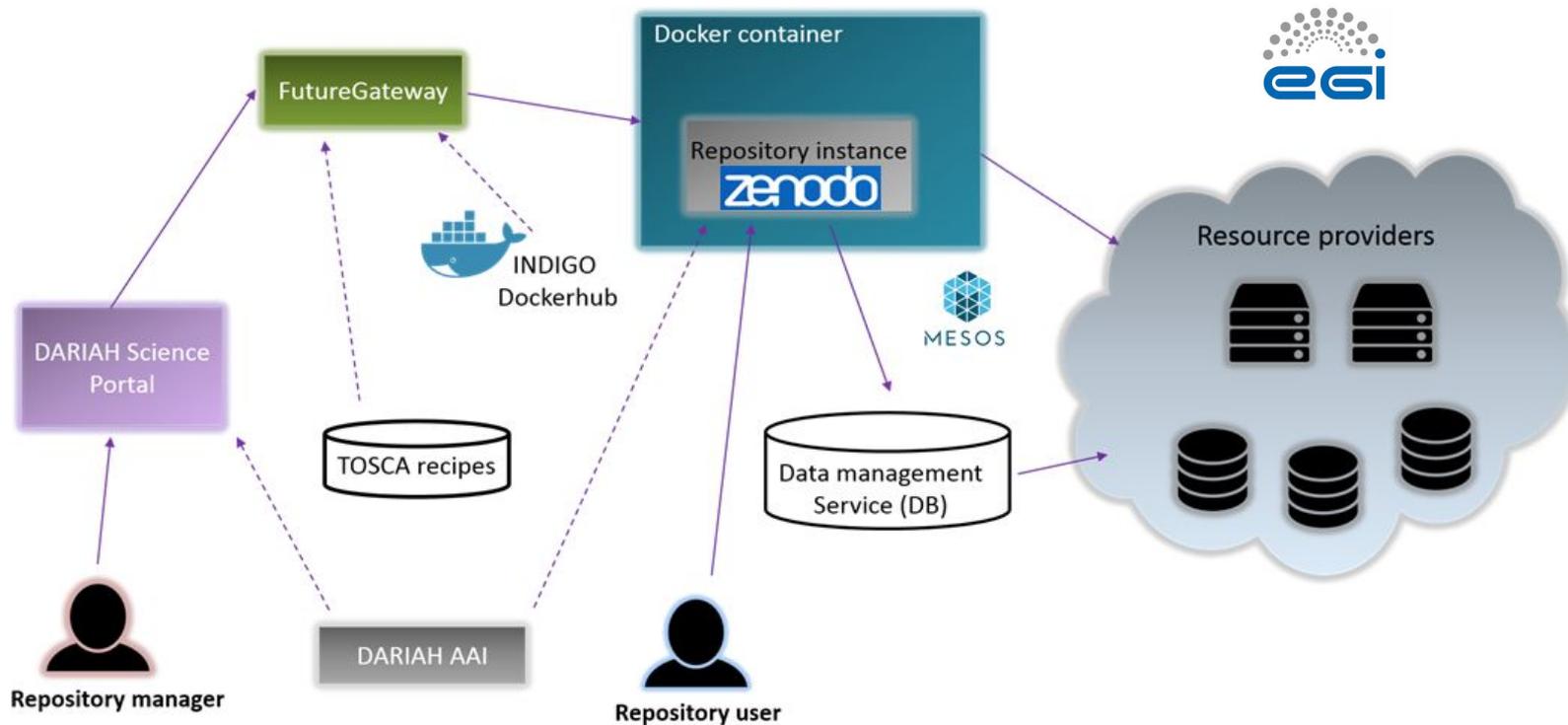
Pokrenuti vlastitu instancu repozitorija temeljenog na Zenodo-u u računalnom oblaku (EGI FedCloud) koristeći Indigo-DataCloud tehnologije

*zbog vremenskog ograničenja i malo zahtjevnijeg postupka pokretanja pokazat ćemo samo demo video

DARIAH repository - scheme



INDIGO - DataCloud



Digitalni repozitorij u Cloudu

Demo

<https://youtu.be/Fk6weIU23pw>

Detalji use case-a dostupni na:

<https://www.indigo-datacloud.eu/communities/champions/davor-davidovic-and-eva-cetinic-rbi-croatia>

Kraj!



INDIGO-DataCloud is funded by the Horizon 2020 Framework Programme grant agreement RIA 653549