

# Europski istraživački prostor i digitalna humanistika

Davor Davidović, Eva Cetinić, Karolj Skala

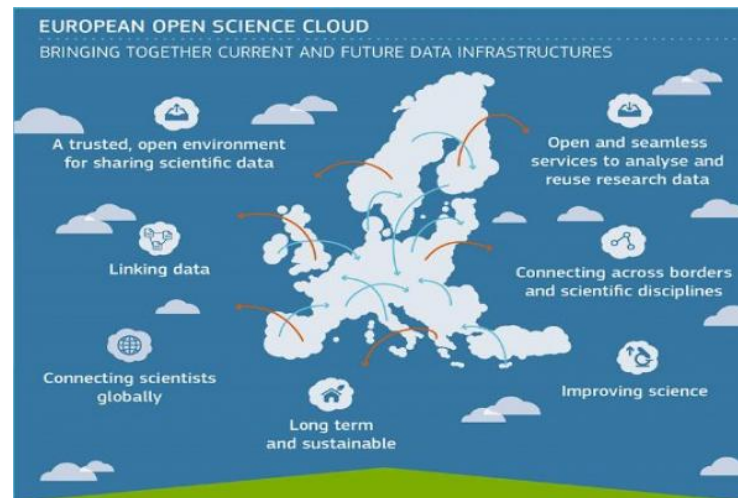
Ruđer Bošković Institute  
DARIAH Competence Centre, EGI-Engage  
INDIGO-DataCloud



# European Open Science Cloud

- „The European Open Science Cloud (EOSC) je vizija **federativne, globalno dostupne, multi-disciplinarne** okoline u kojoj istraživači, inovatori, kompanije i građani mogu međusobno **publicirati, pretraživati** i **koristiti** podatke i druge rezultate za svoja istraživanja, inovacije i edukaciju.”

*(Open Science Policy Platform EOSC wg)*



- cilj:** omogućiti svakom istraživačkom centru, istraživačkom projektu i istraživaču u Europi pristup najmodernijim računalnim, podatkovnim i analitičkim kapacitetima koji su im potrebni kako bi dodatno povećali svoju inovativnost i istraživački potencijal

# Izazovi i prepreke

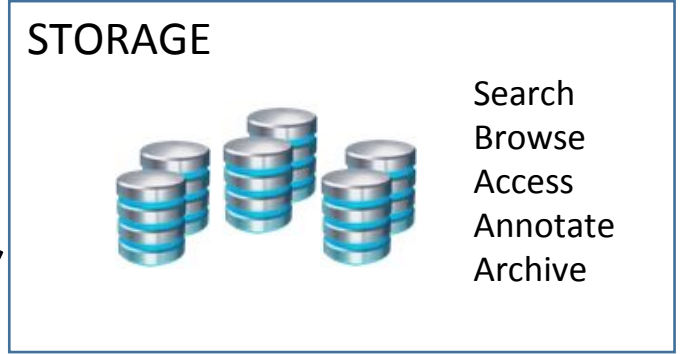
- većina prepreka u realizaciji EOSC ciljeva su više **društvene**, a ne **tehničke** prirode
- glavni tehnički izazov je **kompleksnost podataka** i **analitičkih procedura** unutar različitih znanstvenih disciplina nego sama veličina podataka
- **nedostatak interoperabilnosti** različitih alata, API-a, načina dijeljenja podataka, AAI,...
- **fragmentiranost** – između znanstvenih i ekonomskih domena, nacionalnih i regionalnih granica, upravljačkih modela...
- heterogenost i kratkoročni projektno-orijentirani modeli financiranja (**problem održivosti**)
- zastarjeli sustav financiranja za znanost i inovacije
- podaci dobiveni iz javno financiranih istraživanja nisu uvijek javno dostupni
- nedostatak jasnih prednosti dijeljenja podataka

(izvor: „*Realising the European Open Science Cloud*“, First report and recommendations on the European Open Science Cloud Realising the European Open Science Cloud of the Commission High Level Expert Group)

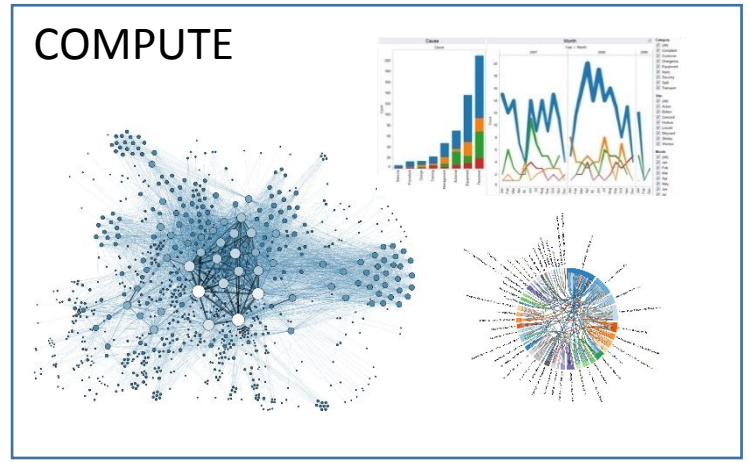
# Digitalna Humanistika



digitalizacija  
pohrana



analiza



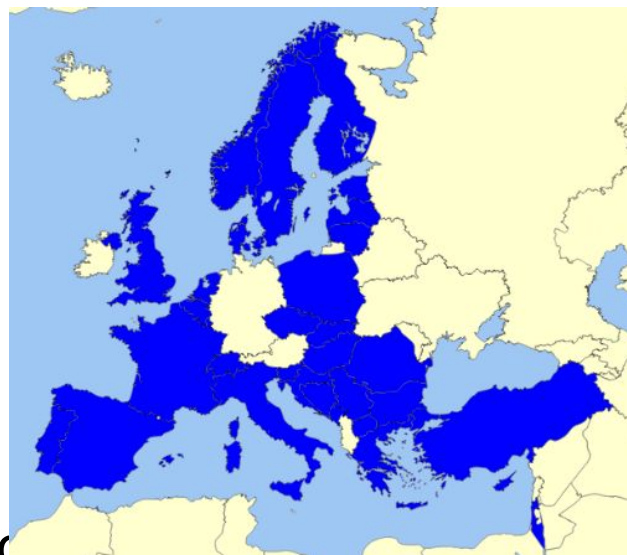
# EGI – European Grid Infrastructure



32 europske zemlje

Federacija:

- **21** pružatelj cloud usluga
- **300** podatkovnih centara koji pružaju računalne resurse i resurse za pohranu podataka



**152.1**  
million

Computing hours  
for HTC

**169.6**  
TB

Storage

**3926**  
GB

RAM memory

**1410**  
vCPUs

Virtual CPUs  
for cloud

## **DARIAH = Digital Research Infrastructure for the Arts and Humanities**

- misija: porast i podrška korištenju digitalnih metoda u istraživačkom i edukativnom radu unutar humanističkih disciplina
- infrastruktura u širem smislu - mreža alata, informacija i ljudi
- DARIAH nije pružatelj e-Infrastrukturnih usluga i ne daje resurse za pohranu i obradu podataka

- Horizon 2020 projekt, 03/2015 - 08/2017
- **cilj:** pojačati korištenje EGI resursa od strane znanstvene zajednice i omogućiti znanstvenicima pristup računalnim resursima, resursima za pohranu podataka i specifičnim aplikacijama
- mreža od **8 centara kompetencije**
- povezivanje e-infrastrukturnih servisa i znanstvenih zajednica kroz:
  - prikupljanje i analizu zahtjeva
  - integraciju domenski specifičnih aplikacija u zrele i dostupne servise
  - razvoj interoperabilnosti među servisima i e-Infrastrukturama

- **osnovni cilj:** potaknuti širu i sveobuhvatniju primjenu EGI servisa u digitalnoj humanistici
- **pod-ciljevi:**
  - proširiti svijest o mogućnostima i prednostima korištenja e-Infrastrukturnih servisa u humanistici
  - povećati broj aplikacija iz domene humanistike koje koriste cloud servise
  - ojačati suradnju između EGI-ja i DARIAH-a

url: [wiki.egi.eu/wiki/Competence\\_centre\\_DARIAH](http://wiki.egi.eu/wiki/Competence_centre_DARIAH)





## Koje su potrebe DARIAH zajednice?

- ne postoji jedinstven odgovor
- heterogena zajednica - različite znanstvene discipline, istraživačke prakse, tipovi podataka, aplikacije...
- prvi korak - anketa o iskustvu korištenja e-Infrastruktura i potrebama korisnika
- resursi kojima korisnici imaju pristup ili žele imati pristup

Digital Repositories	Computational resources (e.g. computer cluster, grid, or cloud)	Authorization / Authentication services	Web-oriented services
33%	22%	22%	22%

# DARIAH izvještaj o potrebama istraživača

- Jenny Oltersdorf, Markus Matoni, Carsten Thiel. “DARIAH Report on researchers’ service needs”, DARIAH, 2016.
- HaS – Humanities at Scale project

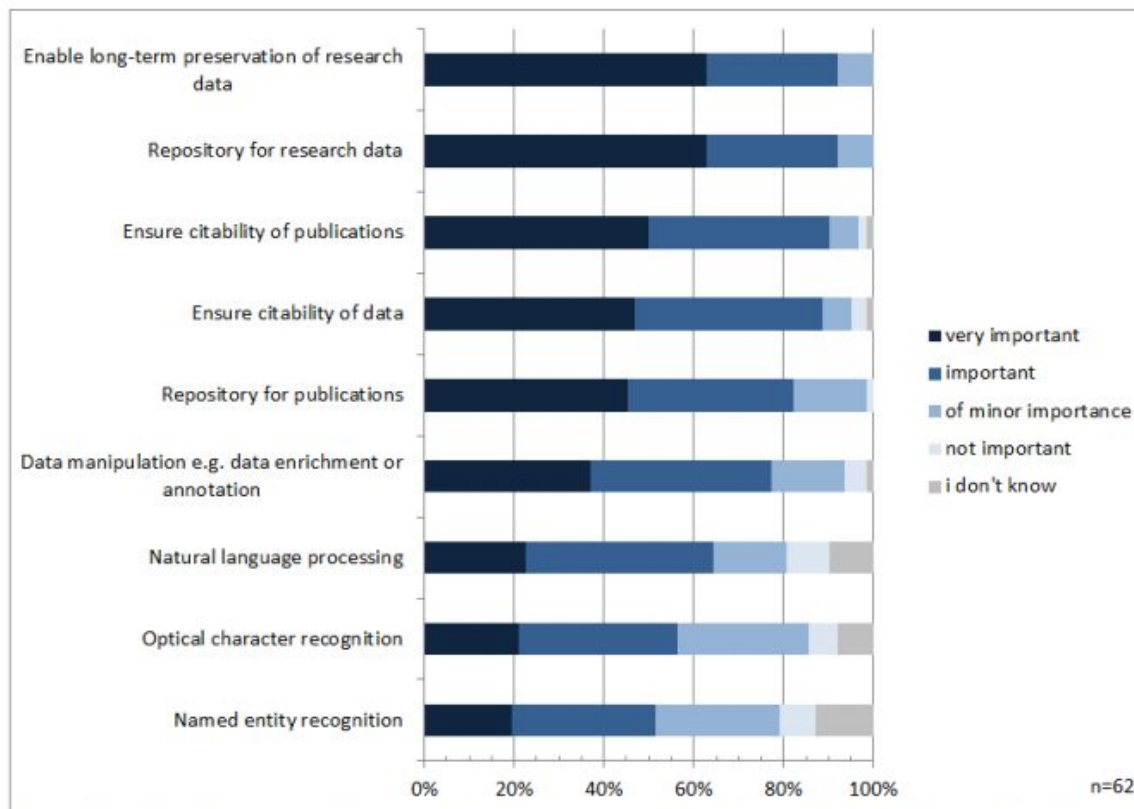
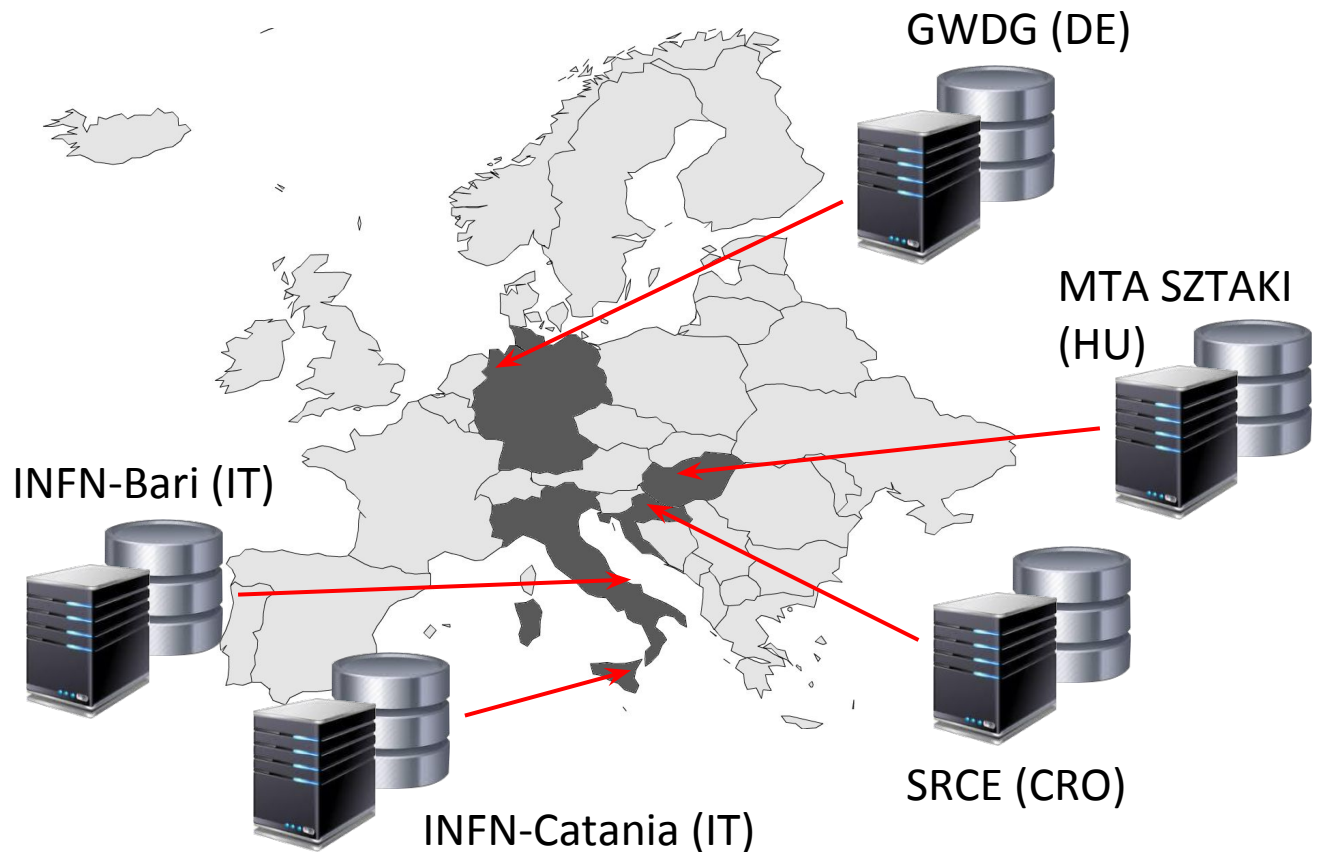


Chart 5: Distribution (in percentage) of research activities referring to all complete replies (n=62). The services and tools are sorted by "very important". Question: How important do you consider infrastructure services or tools to support your research?

# Trenutno dostupni FedCloud resursi (unutar DARIAH zajednice)

- Virtual organization: [vo.dariah.eu](http://vo.dariah.eu)
- EGI-DARIAH SLA: 1/4/2016 – 1/9/2017

VCPU	30
Memory	70 GB
Storage	2 TB



- **DARIAH Science Gateway** (<https://dariah-gateway.lpds.sztaki.hu> )
- Uspostava virtualne organizacije – DARIAH VO (vo.dariah.eu)
- Osigurani FedCloud resursi
- Aplikacije (otvoreni pristup)
  - Semantic Search Engine
    - Simple Semantic Search Engine (SSE)
    - Parallel Semantic Search Engine (PSSE)
  - DBO@Cloud
- Servisi (potrebna autentifikacija)
  - Cloud Access
  - Workflow Development
  - File transfer

# INDIGO-DataCloud



INDIGO - DataCloud

- INDIGO - DataCloud razvija platformu koja objedinjuje skup različitih servisa koji omogućuju lakše korištenje računalnih resursa i resursa za pohranu podataka u cloud okruženju

- **H2020** projekt
- **26** partnera iz **11** europskih zemalja
- **webstranica:** [www.indigo--datacloud.eu](http://www.indigo--datacloud.eu)



- najveći izazov - komunikacija zahtjeva i potreba između znanstvenih zajednica i pružatelja i razvijatelja e-Infrastrukturnih usluga



# Naš cilj – repozitorij u oblaku



INDIGO - DataCloud

- **Što?**
  - personalizirana instanca digitalnog repozitorija (temeljena na open access repozitorskim platformama) u okruženju računalnog oblaka
- **Zašto?**
  - pojednostaviti proces instalacije repozitorija i alokacije potrebnih resursa
- **Za koga?**
  - usmjereno prema potrebama manjih istraživačkih skupina (za rješenja koja nisu institucionalno ovisna)

# Što smo zaključili



- rastući trend potrebe za e-Infrastrukturnim resursima
  - + podrška, alati, edukacija,...
- digitalna humanistika - raznolikost potreba i praksi
- znanstveni portal kao sabirna točka za pristup različitim servisima i resursima
- veliki izazov ostaje kako angažirati potencijalne korisnike i ojačati suradnju
- važnost **povezivanje** različitih struka
- danas je glavni “problem” humanističkih znanosti **pohrana**, sutra će biti **obrada i analiza** podataka!

# Hvala na pažnji



[www.egi.eu](http://www.egi.eu)

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

EGI-Engage is co-funded by the Horizon 2020 Framework Programme of the European Union under grant number 654142

