



Implementacija i korišćenje otvorenih istraživačkih podataka za istraživače OTVORENI ISTRAŽIVAČKI PODACI

Nadica Miljković

Univerzitet u Beogradu – Elektrotehnički fakultet (ETF)

O meni

- Nadica Miljković
 - Docentkinja na ETFu
 - Moja glavna oblast istraživanja je biomedicinsko inženjerstvo ... ali i otvoreni nastavni materijali, rodna ravnopravnost, otvoreni podaci, nauka o podacima ...
 - Imam >10 godina iskustva u Akademiji i >6 godina iskustva u industriji.
-
- Veoma sam zahvalna Milici Ševkušić, Biljani Kosanović, Ljiljani B. Lazarević, Iris Žeželj, Danki Purić i drugima koji su me uveli u svet otvorene nauke i otvorenih podataka.

e-mail: nadica.miljkovic@etf.bg.ac.rs

kabinet 68, Zgrada tehničkih fakulteta

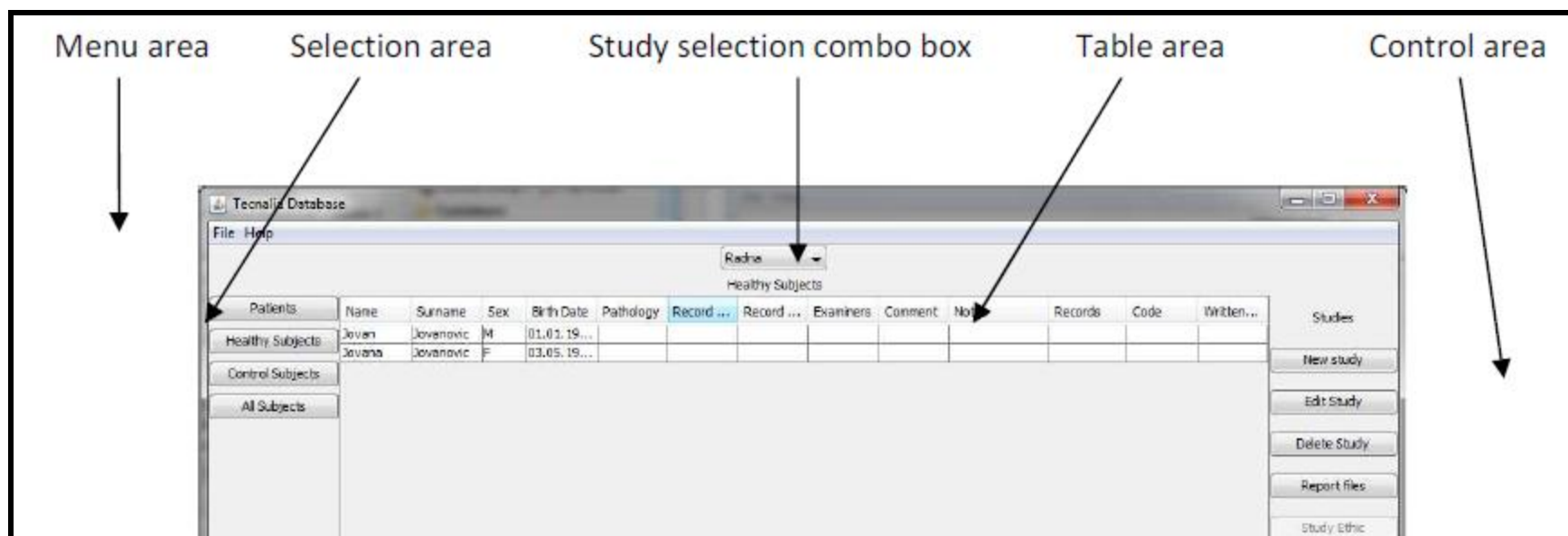


Članica je RDA (Research Data Alliance) i OAW (Open Access Week).

Moje iskustvo sa istraživačkim podacima

- Previše ponavljanja u elektrofiziološkim merenjima. Nedovoljna transparentnost u radu u istoj istraživačkoj instituciji. Nas par je merilo iste signale više puta ...
- Veliki troškovi. Gubitak vremena. Niska interoperabilnost.
- Dodatno, više ispitanika -> bolja analiza. Da li je izvodljivo u kliničkim uslovima?
- Zahvaljujući direktoru u Tecnalia Serbia Ltd. Goranu Bijeliću, realizovana je aplikacija za arhiviranje primarnih istraživačkih podataka sa idejom da jednog dana bude otvorena. Nije zaživela.
- A u nastavi? Predmet Metode analize elektrofizioloških signala (<http://automatika.etf.rs/sr/13e054mas>) i PhysioNet baza (<https://physionet.org/>) za studentkinje i studente ETFa, ali i u istraživanjima.
- Onda smo nedavno podelili podatke za jedan rad u časopisu kao supplementary materijal. Koja greška!

TECNALIA DATABASE USER'S MANUAL



Odabrane stranice iz uputstva za Tecnalia Database aplikaciju iz marta 2012. godine. (Tecnalia Database user's manual for storing and managing measurement data)

Sadržaj

- Cilj prezentacije da se ponude odgovori na pitanja:
 - Šta su otvoreni istraživački podaci (OIP)?
 - Kako nastaje OIP?
 - Kako se pronalazi OIP?
- Nakon ove prezentacija: praktični zadaci za polaznike – pristup otvorenim istraživačkim podacima.



O OTVORENIM ISTRAŽIVAČKIM PODACIMA

Definicija/e

- Otvoreni istraživački podaci su podaci (informacija) koji nastaju kao rezultat istraživačkog procesa i koji su dostupni svima. (moja interpretacija)
- “Open research data refers to the data underpinning scientific research results that has no restrictions on its access, enabling anyone to access it.”
 - Izvor: https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/open-science-monitor/facts-and-figures-open-research-data_en
- “Research data refers to resources which the researcher produces or uses during the research process. It is needed to validate the research and it is meritorious in research if it is published. Open publication of research data is a component area of open science. Data can be stored in Finnish or international repositories. The degree of publicity can be regulated by different licenses (e.g. Creative Commons, CC).”
 - https://uva.libguides.com/open_access/open_research_data

Primarni i sekundarni istraživački podaci

- Podaci mogu biti primarni i sekundarni,
<https://www.statisticshowto.datasciencecentral.com/primary-data-secondary/>
 - Primarni podaci su podaci koji su prikupljeni “iz prve ruke” korišćenjem naučnih metoda (upitnici, eksperimenti, merenja, intervjui) sa jasnim projektnim ciljem.
 - Posebna kategorija primarnih su sirovi (eng. *raw*) podaci,
https://en.wikipedia.org/wiki/Raw_data
 - Sekundarni podaci su podaci koji su prikupljeni iz studija koje su izveli drugi istraživači i sa drugim ciljem i koji se mogu ponovo koristiti (eng. *reuse*).
- Mogu biti i kvalitativni i kvantitativni.
https://en.wikibooks.org/wiki/Statistics/Different_Types_of_Data/P_S.

Primeri istraživačkih podataka

- Istraživački podaci uključuju:
 - rezultate eksperimenata,
 - merenja,
 - observacije,
 - analize upitnika,
 - snimke intervjuja,
 - fotografije i
 - dr.
- Kada se priča o otvorenim istraživačkim podacima, fokus je na njihovoj dostupnosti u digitalnoj formi.
- Izvor: Sarah Jones (iz Digital Curation Centre, University of Glasgow), HORIZON 2020 Open Research Data Pilot“More information on the Open Research Data Pilot in H2020: who should participate and how to comply”, 2014, <https://www.fosteropenscience.eu/content/horizon-2020-open-research-data-pilot>

FAIR podaci

- FAIR je skraćenica koja označava **Findable** (mogućnost pronalaženja), **Accessible** (dostupnost), **Interoperable** (interoperabilnost) i **Reusable** (ponovna upotrebljivost) podataka.
- U okviru radionice pod nazivom “Jointly Designing a Data Fairport” koja se održala u Lajdenu u Holandiji u 2014. godini definisani su FAIR principi koji važe i za ljude i za mašine. Tada je osnovana i FAIR radna grupa.
 - Ovoj radionici su prisustvovali i predstavnici akademije i industrije.
 - Radionica je organizovana jer postoji zajednički interes da se iskoristi maksimalno trud i sredstva koja ulažemo u istraživanja.



Zašto delimo podatke?

- To je uslov za H2020 projekte. Jednostavno rečeno “Public money. Public data.” (po ugledu na kampanju “Public money. Public code.”, <https://publiccode.eu/>)
 - Više na: Sarah Jones (iz Digital Curation Centre, University of Glasgow), HORIZON 2020 Open Research Data Pilot “More information on the Open Research Data Pilot in H2020: who should participate and how to comply”, 2014, <https://www.fosteropenscience.eu/content/horizon-2020-open-research-data-pilot>
- Transparentno istraživanje
- Brži razvoj nauke kroz ponovnu upotrebu postojećih podataka
- Veća citiranost objavljenih istraživanja
- Provera objavljenih rezultata
- Da sačuvamo podatke
- Zato što je etički, pogotovu ako se rade studije na životinjama.

The personal pain of data loss

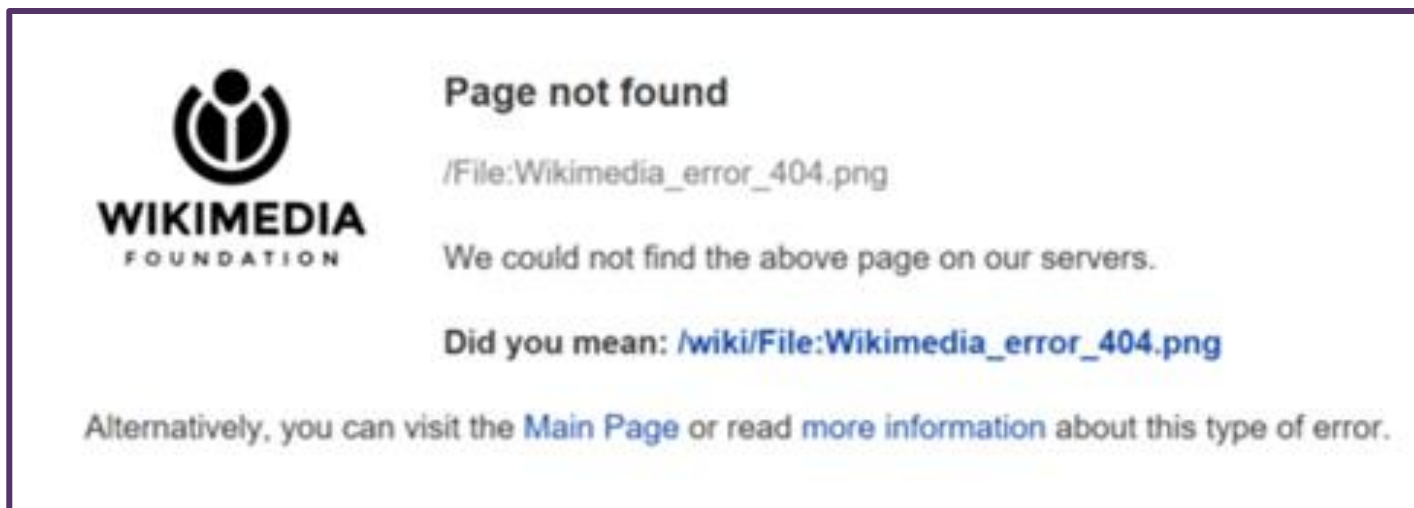


Credit: Dave Hill, CC-BY-NC-SA 2.0
Generic.

<https://www.flickr.com/photos/dmh/650/4031607067/in/gallery-wlef70-72157633022909105/>. Slika takođe u prezentaciji Nielsen, Lars Holm. (2019, April). Zenodo: FAIR data in a generic data repository. Zenodo. <http://doi.org/10.5281/zenodo.2650088>.

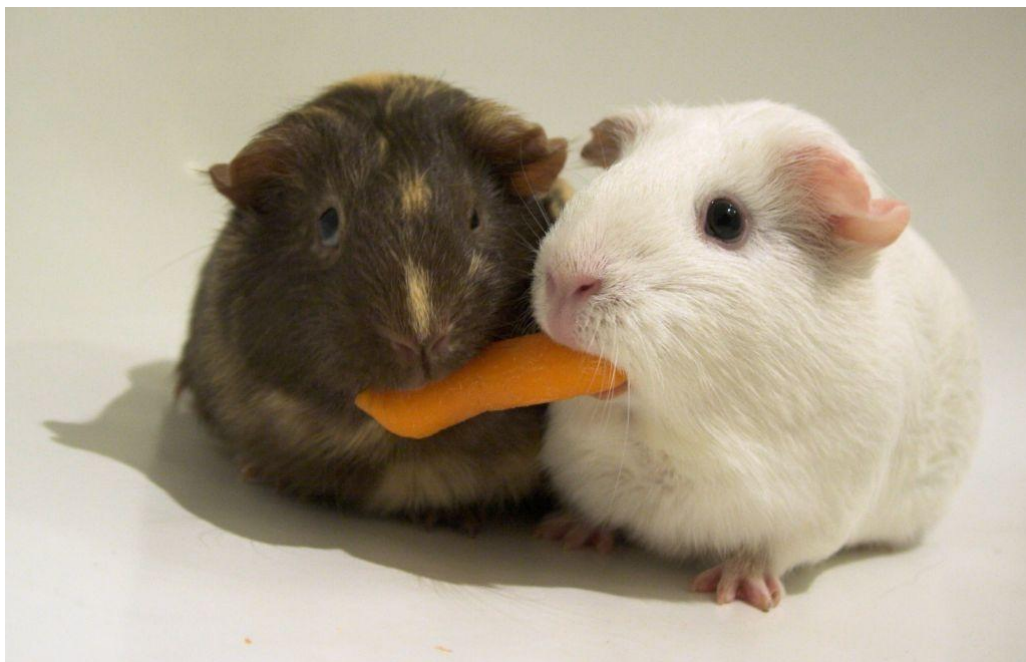
PID je važan za FAIR podatke

- PID (eng. *Persistent Identifier*) je trajna referenca na dokument, fajl, internet stranicu ili druge objekte (https://en.wikipedia.org/wiki/Persistent_identifier).
- “Truo” link (https://en.wikipedia.org/wiki/Link_rot) ili sahranjen link na internetu
 - Postoje studije koje su se bavile prebrojavanjem internet grobova tzv. “404” HTTP poruke greške (https://en.wikipedia.org/wiki/HTTP_404). Podaci se kreću od ~25% do ~50%.



A, da ih ipak ne podelimo?

- Nemamo vremena
- Nema nagrade
- Ako želimo da iskoristimo rezultate u komercijalne svrhe ili u okviru industrije
- Ako na taj način ugrožavamo poverljivost ili bezbednost
- Ako naše deljenje nije u skladu sa GDPRom (eng. *General Data Protection Regulation*, https://en.wikipedia.org/wiki/General_Data_Protection_Regulation)
- Nekada samo privremeno ne želimo da delimo



Šta kaže na to H2020 manual?

- Kompromis je “as open as possible, as closed as necessary”
 - https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Moć podataka

- Postoje mnoga interdisciplinarna istraživanja koja ranije nisu bila moguća.
- Primer: Više noćnih poziva korišćenjem mobilnih telefona će uzrokovati veći broj HIV pozitivnih u određenim područjima Afrike (hvala Draganu Satariću što mi je rekao za ovu studiju)
 - Brdar, Sanja, et al. "Unveiling spatial epidemiology of HIV with mobile phone data." *Scientific reports* 6 (2016): 19342.

Kako se dele istraživački podaci?

- Generalno, u tri koraka (prema Nielsen, Lars Holm. (2019, April). Zenodo: FAIR data in a generic data repository. Zenodo. <http://doi.org/10.5281/zenodo.2650088>)
 - *Upload* / Postaviti na repozitorijum
 - *Describe* / Opisati (uključuje i PID)
 - *Publish* / Objaviti
- Specifično, zavisi od situacije i od repozitorijuma na kom želimo da podelimo podatke.
- Neki postojeći repozitorijumi:
 - Repozitorijumi specifične namene: Genbank (<https://www.ncbi.nlm.nih.gov/genbank/>), Worldwide Protein Data Bank (<https://www.wwpdb.org/>), UniProt (<https://www.uniprot.org/>), Space Physics Data Facility (<https://spdf.gsfc.nasa.gov/>), ...
 - Repozitorijumi opšte namene: Dataverse (<https://dataverse.org/>), FigShare (<https://figshare.com/>), Dryad (<https://datadryad.org/stash>), Mendeley Data (<https://data.mendeley.com/>), Zenodo (<https://zenodo.org/>), DataHub (<https://datahub.io/>), DANS (<https://dans.knaw.nl/en>), ...

Gde naći podatke?

- › Tip 1: Think about the data you need and why you need them.
- › Tip 2: Select the most appropriate resource.
- › Tip 3: Construct your query strategically.
- › Tip 4: Make the repository work for you.
- › Tip 5: Refine your search.
- › Tip 6: Assess data relevance and fitness -for -use.
- › Tip 7: Save your search and data- source details.
- › Tip 8: Look for data services, not just data.
- › Tip 9: Monitor the latest data.
- › Tip 10: Treat sensitive data responsibly.
- › Tip 11: Give back (cite and share data).

Ili primenom platformi i pretraživača: DataCite (<https://datacite.org/>), Google DatasetSearch (<https://toolbox.google.com/datasetsearch>) i druge.

Article Source: [**Eleven quick tips for finding research data**](#)

Gregory K, Khalsa SJ, Michener WK, Psomopoulos FE, de Waard A, et al. (2018) Eleven quick tips for finding research data. PLOS Computational Biology 14(4): e1006038.

<https://doi.org/10.1371/journal.pcbi.1006038>

ZA KRAJ

Kako da se uključite?

- Delite podatke
- Koristite deljene podatke
- Citirajte podatke korišćenjem PID!
- Učlanite se u RDA (Research Data Alliance), ja jesam <https://www.rd-alliance.org/users/nadica-miljkovi%C4%87>.
 - Ali i na <http://openaccessweek.org/>.

Korisni linkovi i literatura

- <https://schoolofdata.org/team/>
- <https://www.fairsfair.eu>
- <http://openaccessweek.org/>
- <https://www.rd-alliance.org/>
- <https://www.pidforum.org/>
- <http://opendata.cern.ch/docs/about-cms>
- <https://datacite.org/>
- Wilkinson, Mark D., et al. "[The FAIR Guiding Principles for scientific data management and stewardship.](#)" *Scientific data* 3 (2016).
- L. Pujol Priego, J. Wareham, "Pistoia alliance", Open science monitor case study – Study, 2019, <https://op.europa.eu/en/publication-detail/-/publication/d94d5843-4547-11e9-a8ed-01aa75ed71a1>.
- Miljković, Nadica, Lazarević B., Ljiljana, Vučkovac, Obrad, & Ševkušić, Milica. (2019, October 23). PSSOH workshop: Open research data practices for researchers (in Serbian: "PSSOH radionica: Implementacija i korišćenje otvorenih istraživačkih podataka za istraživače"). Zenodo. <http://doi.org/10.5281/zenodo.3517034>



PSSOH

Otvoreni istraživački podaci

CONTACT:

Assist. Prof. Nadica Miljković, PhD

University of Belgrade – School of Electrical Engineering

e-mail: nadica.miljkovic@etf.bg.ac.rs, URL: <https://bit.ly/2pvosx0>

