

Šta je upravljanje istraživačkim podacima? Zašto je to uopšte važno?

Obrad Vučkovic

Biblioteka Instituta za nuklearne nauke „Vinča“

Univerzitet u Beogradu

obrad.vuckovic@vin.bg.ac.rs

2. novembar 2019, PSSOH Workshop

Šta je RDM?

Upravljanje istraživačkim podacima se odnosi na niz aktivnosti kako bi se na najefikasniji način prikupili i organizovali podaci, i pripremili za buduću upotrebu i deljenje sa drugima.

Upravljanje istraživačkim podacima = **dobra naučna praksa**

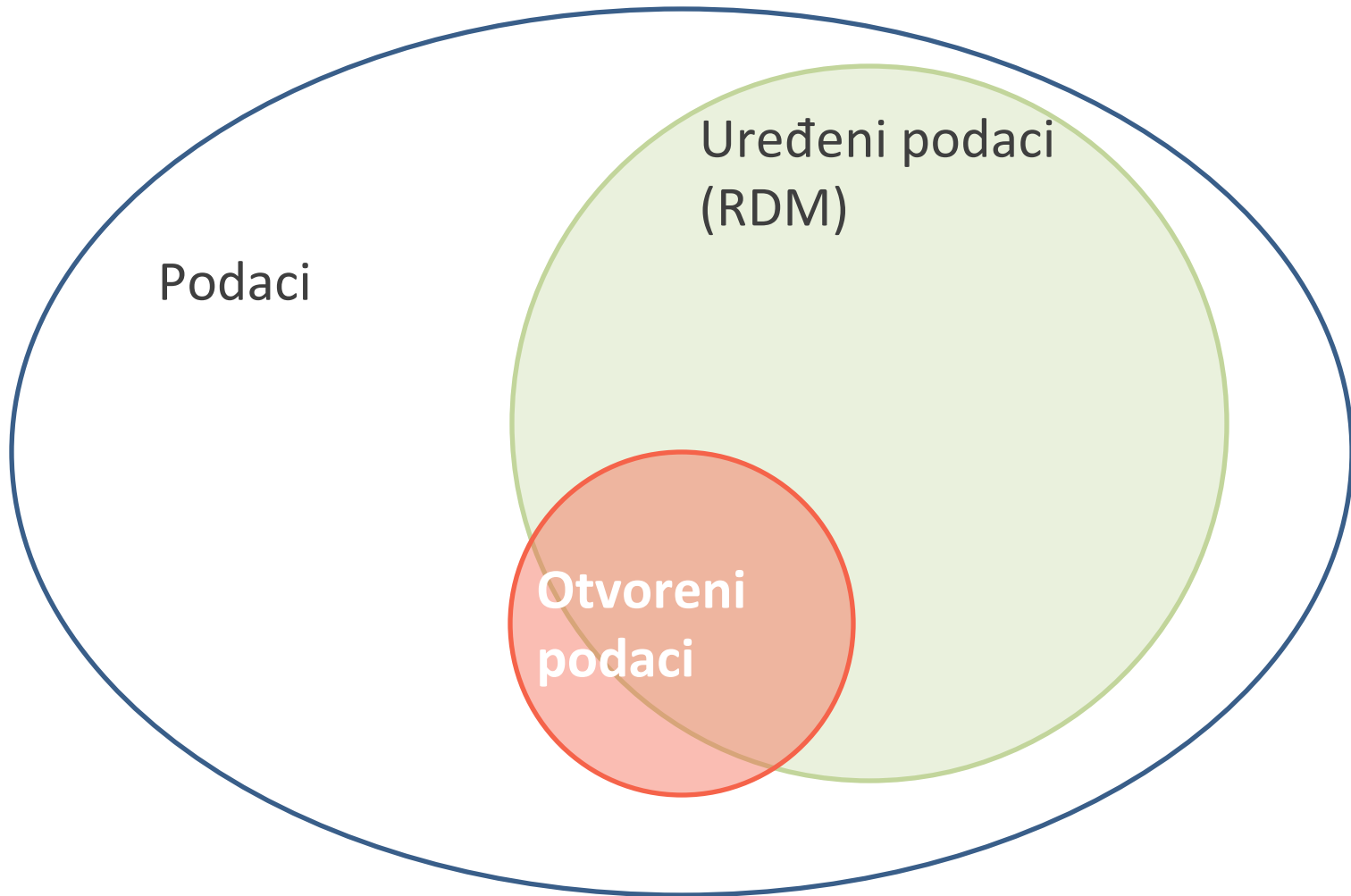
Research data management is *“the active management and appraisal of data over the lifecycle of scholarly and scientific interest.”*

Digital Curation Center, UK

Zašto upravljati podacima?

- Omogućava se provera i ponavljanje eksperimenata od strane drugih istraživača
- Zahtevi naučnih institucija, izdavača i finansijera istraživanja
- Sprečavaju se gubici dragocenih istraživačkih podataka
- Bolji kvalitet istraživanja, sprečavaju se greške
- Može dovesti do novih vidova saradnji

Otvoreni i uređeni podaci



FAIR principi



Findable

Perzistentni identifikatori
Metapodaci

Interoperable



Dogovoreni formati
Kontrolisani rečnici



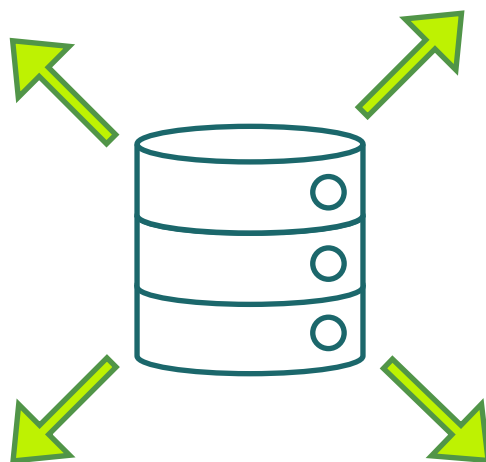
Accessable

Standardizovani
komunikacioni protokoli

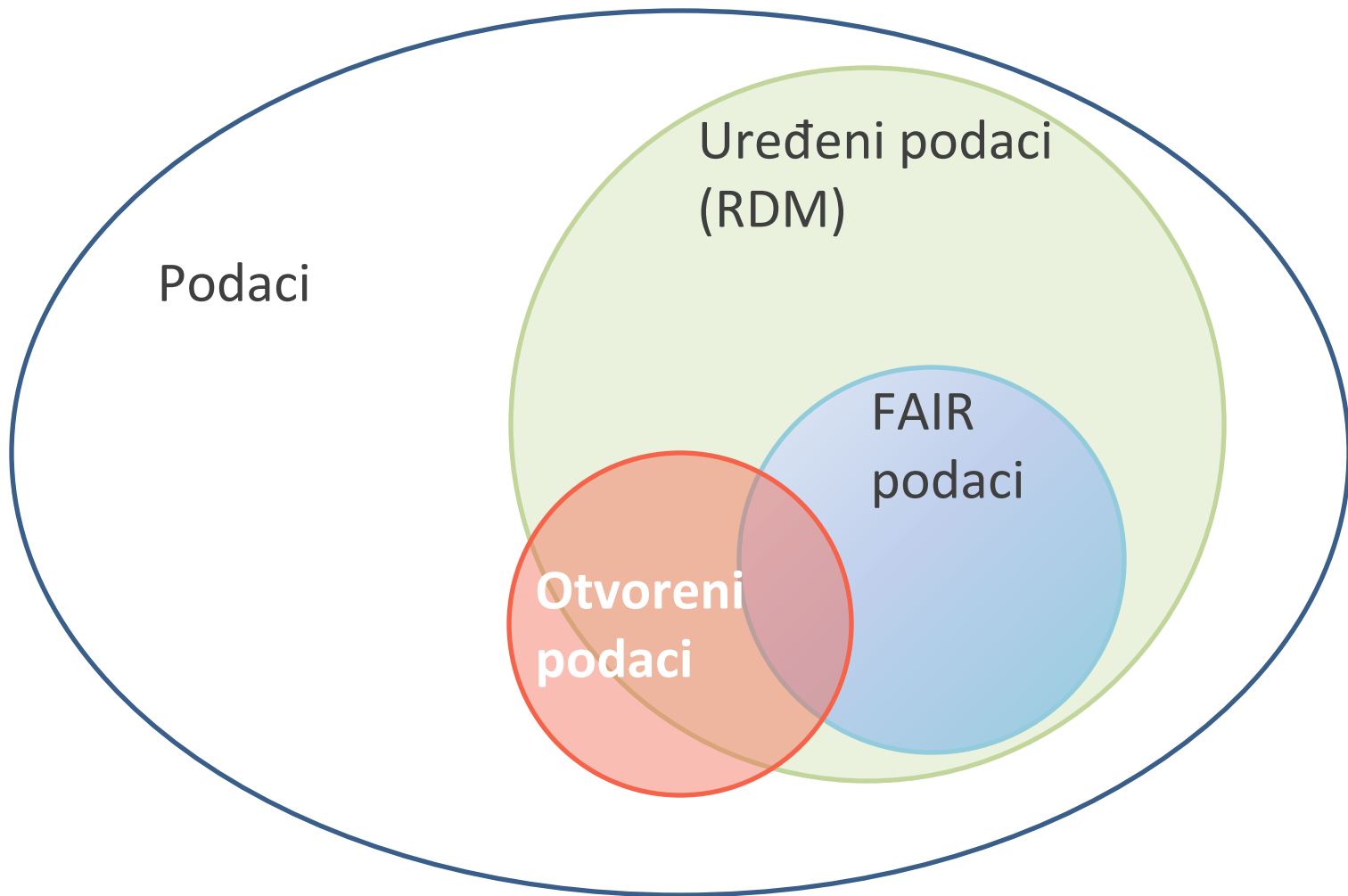
Reusable



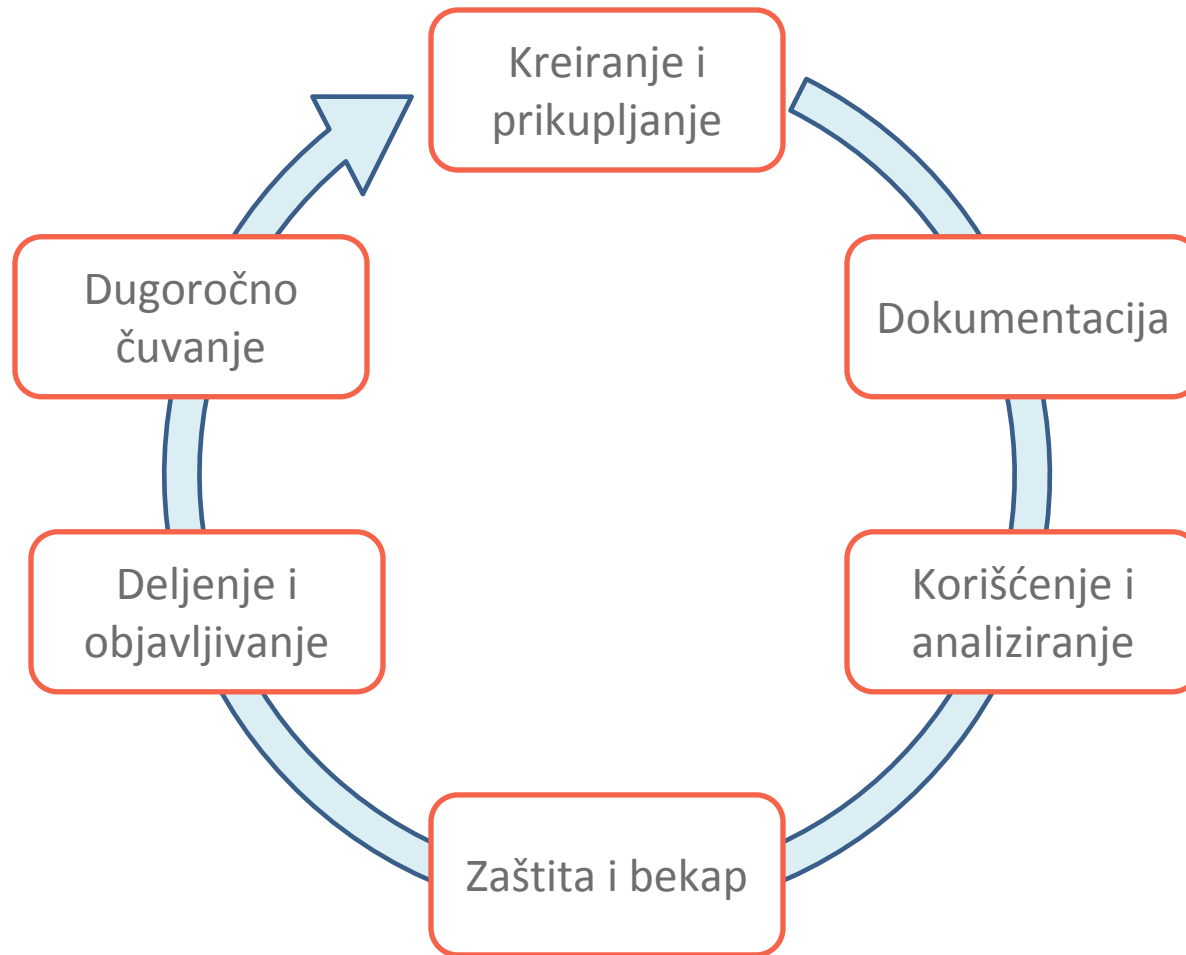
Licence
Poreklo (*Provenance*)



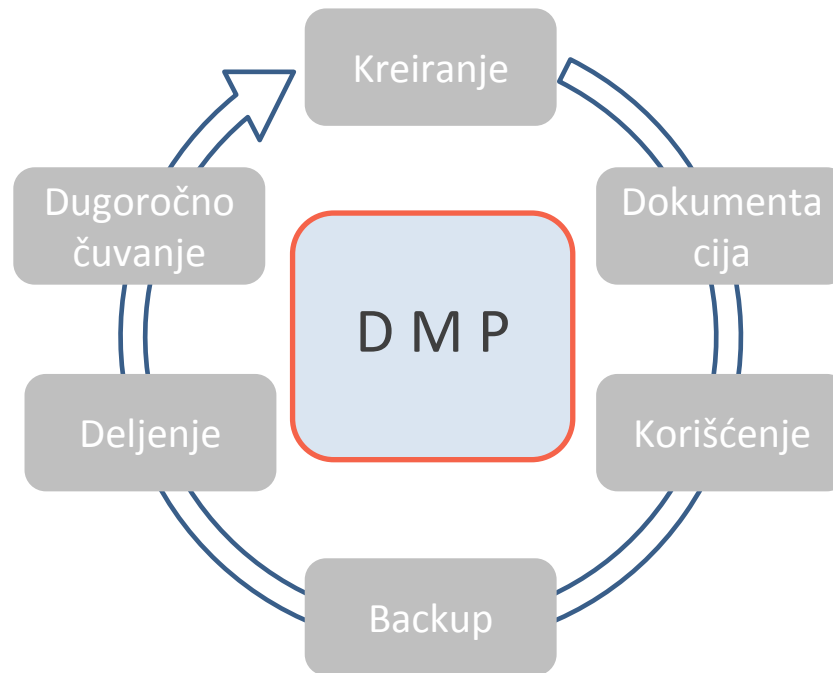
FAIR



Data Lifecycle



Data Management Plan



Dokument koji služi kao okvir svim učesnicima na projektu kako da upravljaju i čuvaju podatke.

Donosi se na početku projekta i treba da su svi učesnici upoznati.

Data Management Plan

Obaveza većine finansijera
(Evropska komisija, Wellcome
trust, NIH, Bill & Melinda Gates
foundation, ...)

H2020 počeo sa DMP kao pilot
projekat; od 2018. je obavezan

Alati za izradu DMP-a:

- DMPOnline (DCC) i
- DMP Tool (DataONE).



Image by [aymane jidi](#) from [Pixabay](#)

Kreiranje i prikupljanje



Image by [The National Archives UK](#) from [Wikimedia Commons](#)

Odabrati odgovarajući format

- otvoreni i *lossless* formati - održivi i bolji za dugoročno čuvanje
npr. XML, CSV, PDF/A, TIFF, WAV
- vlasnički i/ili kompresovani – često u širokoj upotrebi, nisu za dugoročno čuvanje
npr. DOC, XLS, JPG, MP3

Jedan format za analizu, za arhiviranje - trajni format!

Kreiranje i prikupljanje

Organizovanje po folderima

- Imenovati folder po oblasti rada, a ne po istraživaču
- Hijerarhijska struktura – od široke ka užoj oblasti
- Odvajati radne od urađenih (gotovih) foldera

Imenovanje fajlova

- Odrediti zajednički rečnik
- Odvajanje reči: velika slova, crtice (_ ili -); izbegavati dijakritičke znake i prazninu
- Red reči: fajlovi sa istom temom su grupisani i lako se nalaze

Verzije

Svaka bitnija promena na fajlu treba da dobije novu verziju (npr. v2 ili v3_2)
Poslednji fajl dobija verziju „**final**“.

Dokumentacija

```
1 /*
2  * This line basically imports the "stdio" header file, part of
3  * the standard library. It provides input and output functionality
4  * to the program.
5  */
6 #include <stdio.h>
7
8 /*
9  * Function (method) declaration. This outputs "Hello, world\n" to
10 * standard output when invoked.
11 */
12 void sayHello(void) {
13     // printf() in C outputs the specified text (with optional
14     // formatting options) when invoked.
15     printf("Hello, world!\n");
16 }
17
18 /*
19 * This line basically imports the "stdio" header file, part of
20 * the standard library. It provides input and output functionality
21 * to the program.
22 */
23 #include <stdio.h>
24
25 /*
26 * Function (method) declaration. This outputs "Hello, world\n" to
27 * standard output when invoked.
28 */
29 void sayHello(void) {
30     // printf() in C outputs the specified text (with optional
31     // formatting options) when invoked.
32     printf("Hello, world!\n");
33 }
34 }
```



Image by [TeroVesalainen](#) from [Pixabay](#)

Dokumentacija

sva propratna građa koja omogućava pravilnu upotrebu podataka

- upitnici i pitanja za intervju,
- laboratorijske sveske,
- programski kod,
- informacije o podešavanjima i kalibraciji uređaja,
- *codebooks*,
- dokumentacija na nivou fajla (npr. kolone u spreadsheet fajlovima)

Metapodaci

- Standardizovani
- Strukturisani
- Čitljivi za ljude i mašine

Zaštita i *backup*



Image by [John](#) from [Wikimedia Commons](#)

Zaštita i *backup*

Gde skladištiti podatke?

- Laptop, fleš drajev (mogu da se pokvare ili budu ukradeni/izgubljeni)
- Mrežni server
- *Cloud storage* (Dropbox, Google Drive, etc.)

Pravilo 3-2-1 - 3 kopije na 2 različita medijuma i 1 *offsite* (server ili cloud)

Backup i dugoročno čuvanje nije isto!!!



Dropbox



Zaštita i *backup*

Kontrola pristupa

Lozinke

Osetljivi podaci, podaci o ličnosti, poslovne tajne, ...

LastPass – brauzer dodatak (*add-on*) za čuvanje lozinki

Enkripcija













Proces promene podataka u nečitljivi kod.






- BitLocker (za Windows)
- File Vault (za Mac)



Image by [Pete Linforth](#) from [Pixabay](#)

Deljenje

CREATIVE COMMONS LICENSES		 COPY & PUBLISH	 ATTRIBUTION REQUIRED	 COMMERCIAL USE	 MODIFY & ADAPT	 CHANGE LICENSE
	PUBLIC DOMAIN	✓	✗	✓	✓	✓
	CC BY	✓	✓	✓	✓	✓
	CC BY-SA	✓	✓	✓	✓	✗
	CC BY-ND	✓	✓	✓	✗	✓
	CC BY-NC	✓	✓	✗	✓	✓
	CC BY-NC-SA	✓	✓	✗	✓	✗
	CC BY-NC-ND	✓	✓	✗	✗	✓

				
You can redistribute (copy, publish, display, communicate, etc.)	You have to attribute the original work	You can use the work commercially	You can modify and adapt the original work	You can choose license type for your adaptations of the work.

Deljenje

Otvorena nauka => Otvoreni podaci

Ipak, treba obratiti pažnju na:

- zaštitu podataka o ličnosti
- intelektualnu svojinu, zaštitu konkurentnosti, ...

H2020: **'as open as possible, as closed as necessary'**

Embargo period

Dugoročno čuvanje

Kriterijumi za izbor repozitorijuma:

- Domain specific > general repository
- Koje formate prihvataju, *Open/Restricted Access*
- Perzistentni i jedinstveni identifikatori
- Sertifikati (DSA – Data Seal of Approval, ISO 16363)

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

I za kraj...

Curry, A. (2011). Rescue of Old Data Offers Lesson for Particle Physicists. *Science*, 331(6018), 694–695.
<https://doi.org/10.1126/science.331.6018.694>

Hvala na pažnji

Pitanja?

Obrad Vučkovic

Biblioteka Instituta za nuklearne nauke „Vinča“

Univerzitet u Beogradu

obrad.vuckovic@vin.bg.ac.rs