



Univerza v Novem mestu  
*University of Novo mesto*

# Uvod v ravnanje z raziskovalnimi podatki

Delavnica, 13. november 2023

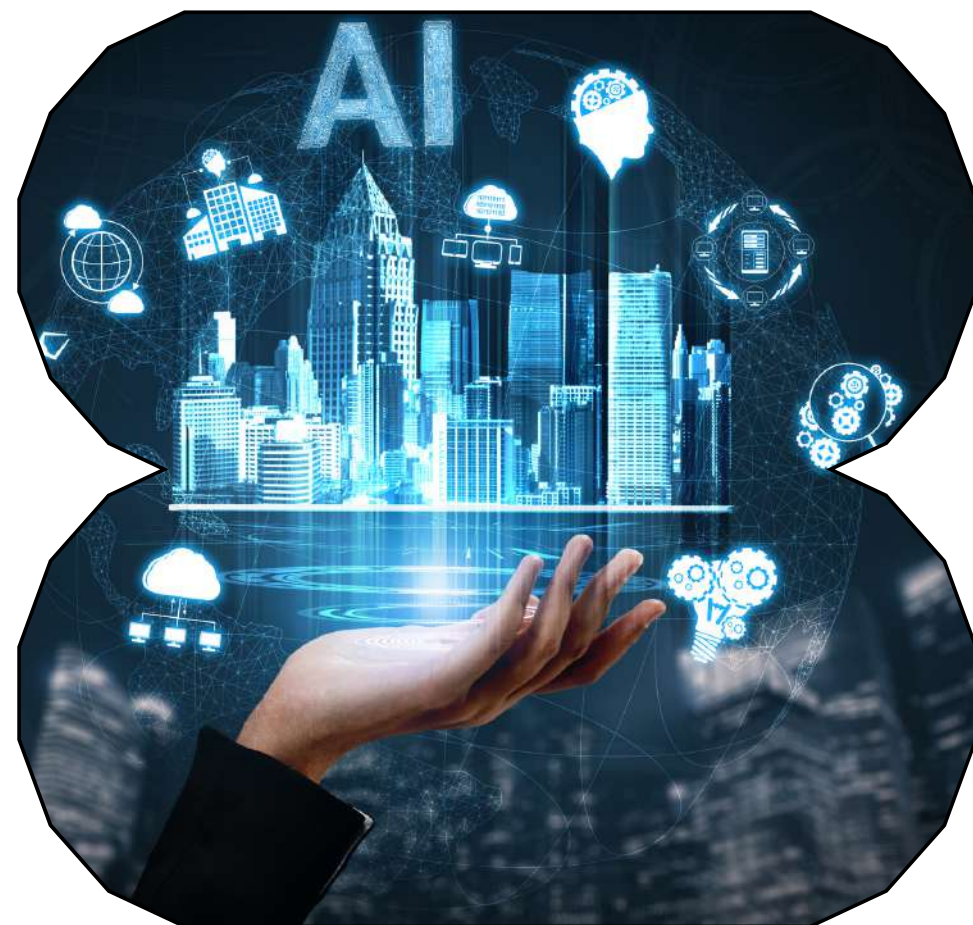




# Kaj je odprta znanost?

“Odprta znanost pomeni nov pristop k znanstvenemu procesu, ki temelji na skupnih prizadevanjih in **novih načinih razširjanja znanja** ter izboljšuje **dostopnost** do rezultatov raziskav in njihovo ponovno uporabo s pomočjo digitalnih tehnologij in novih orodij za sodelovanje.”

(PRIPOROČILO KOMISIJE (EU) 2018/790 z dne 25. aprila 2018 o dostopu do znanstvenih informacij in njihovem arhiviranju)



# Kaj so raz. podatki?

“Raziskovalni podatki predstavljajo **osnovno podlago** za znanstveno raziskovanje in z analizo omogočajo izpeljavo teoretično ali uporabno naravnanih zaključkov. Odprto dostopni podatki morajo biti **pripravljeni** v takšni obliki, da se uporabniki, ki so jim namenjeni, z njimi lahko **seznanijo** in jih **uporabljajo** za različne druge namene.”

(Štebe in dr., 2015) Štebe, Janez, Bezjak, Sonja, Vipavc Brvar, Irena (2015).

Priprava raziskovalnih podatkov za odprti dostop : priročnik za raziskovalce. URN:NBN:SI:DOC-06SLBVXX from <http://www.dlib.si>



## Zakaj uvajamo odprto znanost?

### Zahteve odločevalcev:

- Evropska komisija
- Znanstvene revije
- Raziskovalne organizacije
- Izobraževalne ustanove
- Drugi financerji raziskav

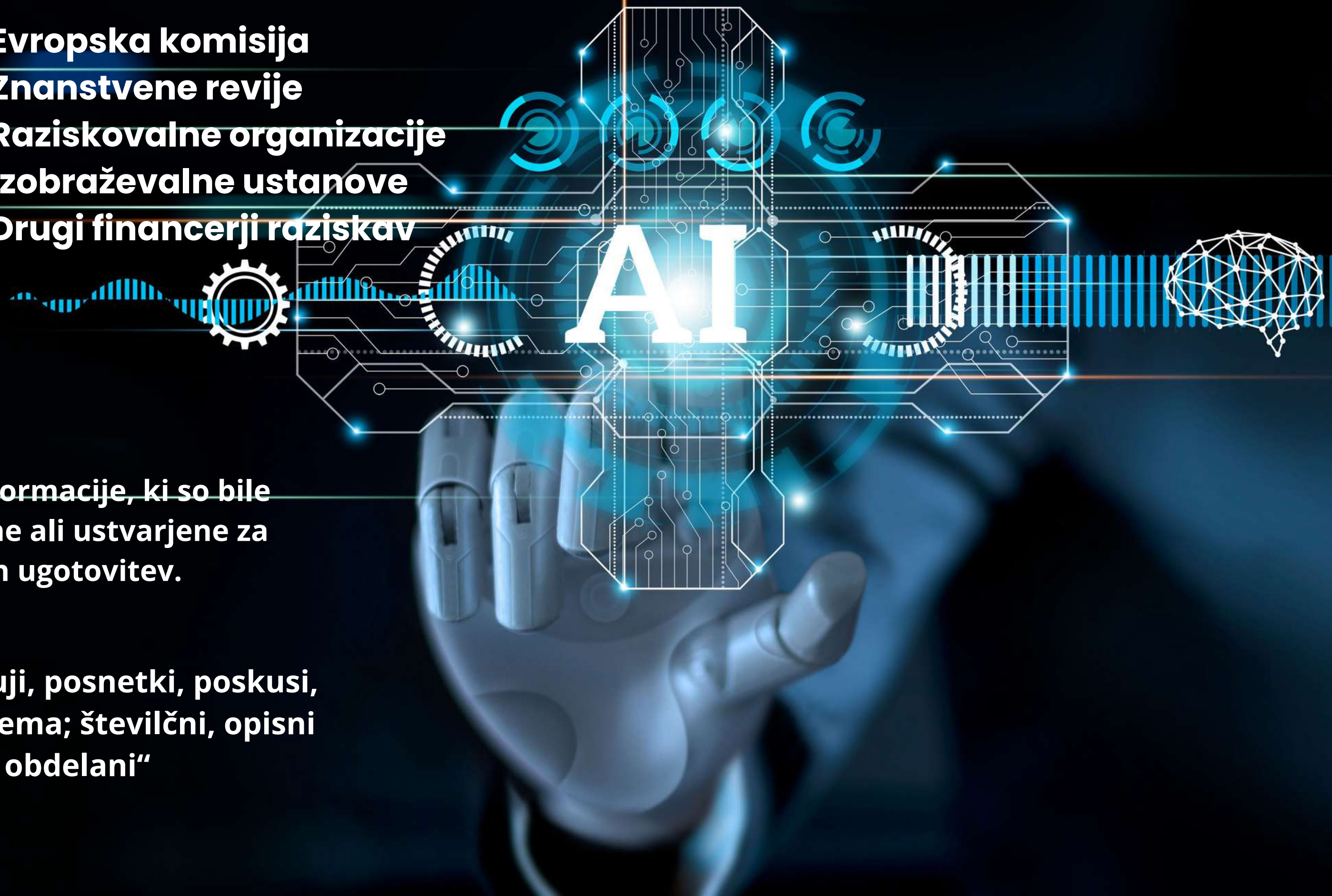
## Kaj so raziskovalni podatki?

Raziskovalni podatki so vse informacije, ki so bile zbrane, opazovane, proizvedene ali ustvarjene za potrditev izvirnih raziskovalnih ugotovitev.

(Univerza v Leedsu, Knjižnica)

„dejstva, opazovanja, intervjuji, posnetki, poskusi, simulacije in programska oprema; številčni, opisni in vizualni; surovi, očiščeni in obdelani“

(Van Berchum & Grootveld, 2017).



## Tipi podatkov

- Numerični (tabele, rezultati meritev, števci, podatki iz senzorjev ...).
- Tekstovni (knjige, zapiski, ankete ...).
- Avdio vizualni podatki (slike, zvočni zapisi, video, animacija).
- Prostorski.
- Specifični glede na znanstveno disciplino.
- Specifični glede na merilni instrument ali napravo.





## Kaj so vaši raz. podatki?

Podatki:

- Surovi/Očiščeni
- Nekontrolirano zbrani
- Eksperimentalni
- Analitični
- Strukturirani/Nestrukturirani
- Stisnjeni/nestisnjeni
- Odprti/Omejeno dostopni
- Nepovezani/Povezani
- Občutljivi
- Metapodatki
- Velepodatki

# Občutljivi podatki

- Vsebujejo potencialno **identifikacijske informacije** o udeležencu raziskave:
  - podatke o **ljudeh** ali druge vrste občutljivih podatkov,
  - ki jih treba **spremeniti**, preden jih delimo z javnostjo.
- Zakaj? Zaradi **zaščite zaupnosti udeležencev** ali iz drugih pomembnih razlogov (npr. nacionalna varnost, varstvo intelektualne lastnine).

```
dropdown-menu"), d=b.data("target"); if (b[0]) { ...  
st a"), f=a.Event("hide.bs.tab",{relatedTarget:b[0]}), g=a.Event("show.bs.tab",{relatedTarget:c[0]}),  
FaultPrevented(){var h=a(d);this.activate(b.closest("li"),c),this.activate(h,h.parent(),function()  
rigger({type:"shown.bs.tab",relatedTarget:e[0]}))}}},c.prototype.activate=function(b,d,e){func  
u > .active").removeClass("active").end().find('[data-toggle="tab"]').attr("aria-expanded",!1),l  
ia-expanded",!0),h?(b[0].offsetWidth,b.addClass("in")):b.removeClass("fade"),b.parent(".dropdo  
(.find('[data-toggle="tab"]').attr("aria-expanded",!0),e&&e())var g=d.find("> .active"),h=e&&  
le")||!d.find("> .fade").length);g.length&&h?g.one("bsTransitionEnd",f).emulateTransitionEnd  
;var d=a.fn.tab;a.fn.tab=b,a.fn.tab.Constructor=c,a.fn.tab.noConflict=function(){return a.fn.t  
"show"));a(document).on("click.bs.tab.data-api",[data-toggle="tab"],e).on("click.bs.tab.data  
se strict";function b(b){return this.each(function(){var d=a(this),e=d.data("bs.affix"),f="ob  
-typeof b&&e[b]()}}var c=function(b,d){this.options=a.extend({},c.DEFAULTS,d),this.$target=a  
!,a.proxy(this.checkPosition,this)).on("click.bs.affix.data-api",a.proxy(this.checkPositionWi  
null,this.pinnedOffset=null,this.checkPosition());c.VERSION="3.3.7",c.RESET="affix affix-top  
State=function(a,b,c,d){var e=this.$target.scrollTop(),f=this.$element.offset(),g=this.$targ  
"bottom"==this.affixed)return null!=c?!(e+this.unpin<=f.top)&&"bottom":!(e+g<=a-d)&&"bottom"  
!=c&&e<=c?"top":null!=d&&i+j>a-d&&"bottom"},c.prototype.getPinnedOffset=function(){if(this  
:RESET).addClass("affix");var a=this.$target.scrollTop(),b=this.$element.offset(),c=br  
WithEventLoop=function(){setTimeout(a.proxy(this.checkPosition,this),100);},a.proxy(this.c  
nt.height(),d=this.options.offset)};return this}
```



# Posredni in neposredni identifikatorji oseb



- **Neposredni identifikatorji:** kažejo neposredno na posameznika in so **običajno odstranjeni** iz podatkovnih naborov.

To so: ime, začetnice, poštni naslov, telefonsko številko, e-poštni naslov, edinstvene identifikacijske številke, kot so EMŠO, davčna številka ali številke voznškega dovoljenja, identifikatorji vozil, identifikatorji medicinskih pripomočkov, spletni ali IP naslovi, biometrični podatki, fotografije osebe, zvočni posnetki, imena sorodnikov, datumi, specifični za posameznika, na primer datum rojstva, poroke itd.

- **Posredni identifikatorji:** sami po sebi neškodljivi, vendar lahko kažejo na posameznika v kombinaciji z drugimi podatki. Priporočeno je, da nabore podatkov, ki vsebujejo **tri ali več posrednih** identifikatorjev, pregleda neodvisni raziskovalec ali odbor za etiko, da oceni tveganje identifikacije. **Vse posredne informacije, ki niso potrebne za analizo, je treba odstraniti** ali predložiti v **agregirani obliki** (na primer razponi letnih dohodkov namesto natančnih številok. To so: kraj zdravljenja ali ime zdravnika, spol, redko bolezen ali zdravljenje, občutljive podatke, kot je uporaba prepovedanih drog ali druga "tvegana vedenja", kraj rojstva, socialno-ekonomske podatke, kot so delovno mesto, poklic, letni dohodek, izobrazba, itd., splošni geografski kazalniki, kot so poštna številka prebivališča, sestava gospodinjstva in družine, etnična pripadnost, leto ali starost rojstva, dobesedni odgovori ali prepisi.



## Priporočeni formati datotek

- Slike: JPEG, JPG-2000, PNG, TIFF
- Tekst: plain text (TXT), HTML, XML, RDF, PDF/A
- Avdio: AIFF, WAVE, MP3
- Video: MPEG4, JPEG-2000
- Stiskanje in združevanje datotek: TAR, GZIP, ZIP
- Podatkovne baze: RDF, XML, JSON, JSON-LD ali CSV namesto binarnih formatov aplikacij



## **Tabelarični podatki**

- Ne dajajte v eno preglednico več kot eno tabelo.
- V prvi vrstici preglednice naj bodo definirane spremenljivke, da jih lahko drugi razumejo.
- Grafe dodajte v nove zavihke. Ne dajajte jih v isti zavihek.
- Pretvorite tabelarične podatke iz delno odprtih formatov (npr. XLS) v odprte formate (CSV ali TSV).



## Metapodatki (za opis)

- Vsebinski
- Strukturni
- Skrbniški
- Bere jih lahko človek (angl. human readable)
- Strojno berljivi (angl. machine readable)
- Strojno izvedljivi (angl. machine actionable)

Primeri metapodatkovnih formatov: DC, COMARC, PREMIS, DCAT, FOAF



# Metapodatkovni standardi – osnovni

## *Dublin Core*

- Title
- Creator
- Subject
- Description
- Publisher
- Contributor
- Date
- Type
- Format
- Identifier
- Source
- Language
- Relation
- Coverage
- Rights

## *DataCite*

- Title
- Creator
- Publisher
- Identifer
- Publication Year
- Resource Type
- Subject
- Contributor
- Date
- Related identifier
- Description
- Geolocation
- Language
- *Alternate identifier*
- *Size*
- *Format*
- *Version*
- *Rights*
- *Funding Reference*

## *EDMI*

- Name
- Description
- Identifier
- url
- Creator
- Date Created
- license
- Data Standard
- Date Modified
- Access URL
- Access Interface
- Structure
- Included In
- Measurement Technique
- Keywords
- Variable Measured
- Format
- Scientific Type
- Includes
- Content Type
- Size
- Authentications
- *Version*
- *Metric*
- *Same as*
- *Spatial Coverage*
- *Temporal coverage*
- *Citation*
- *Reference citation compression*



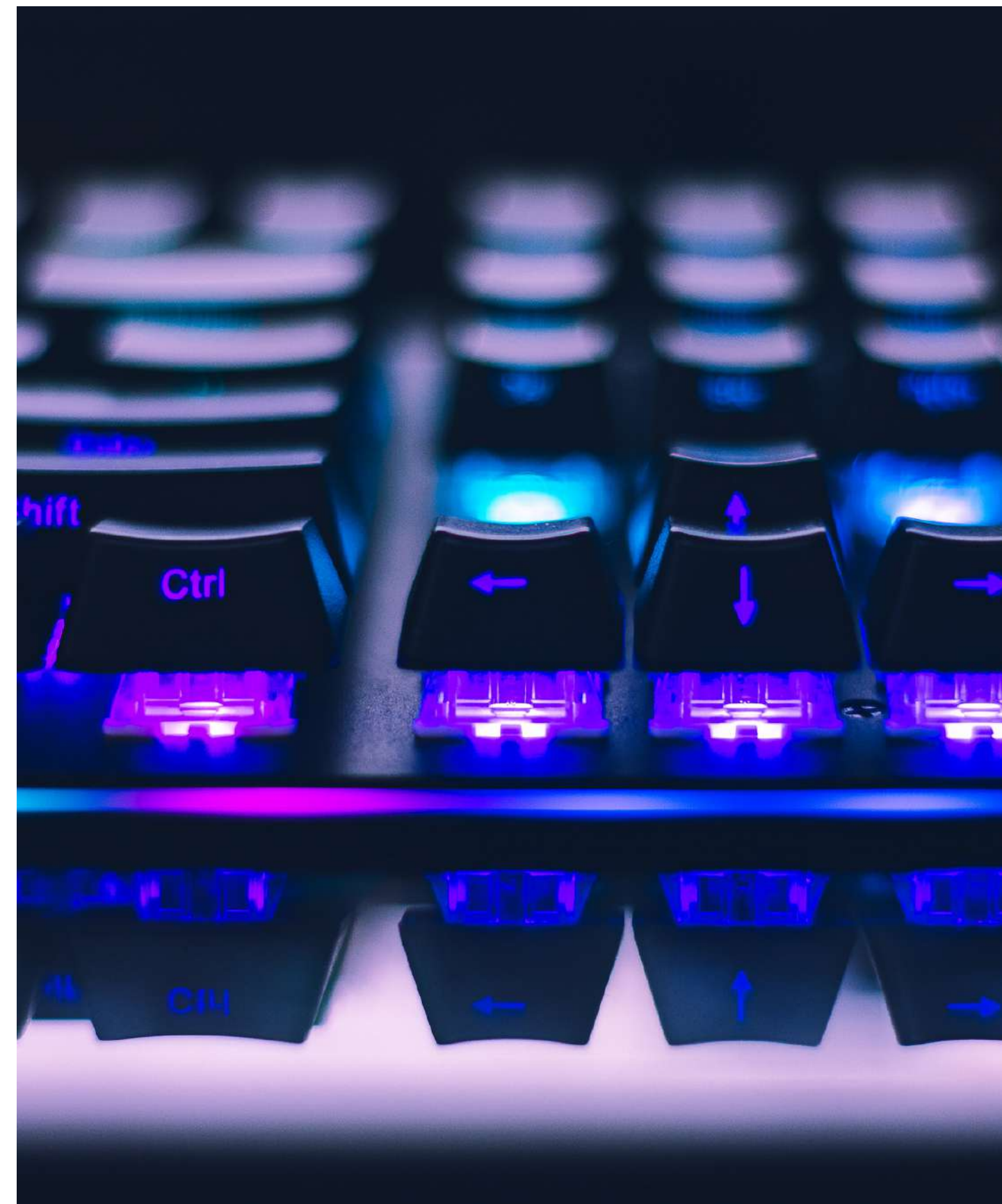
# Ravnanje z raziskovalnimi podatki



CESSDA Training Team (2017 - 2022). *CESSDA Data Management Expert Guide*. Bergen, Norway: CESSDA ERIC. Retrieved from <https://dmeg.cessda.eu/>



„Življenjski krog raziskovalnih podatkov je model, ki ponazarja faze ravnanja s podatki in opiše pretok podatkov od začetka do konca raziskovalnega projekta.“  
(Princeton Research Data Service)





# Načrt ravnanja z raziskovalnimi podatki

„Z njim opredelimo **ključne aktivnosti in strategije**, ki pomagajo zagotoviti, da so podatki visoke kakovosti, varni, trajnostni in – kadar je mogoče – dostopni in ponovno uporabni.“

(OpenAire Research Data Management Handbook)

Je **dokument**, v katerem opišemo, **kakšne podatke bomo zbirali**, kako bomo z njimi ravnali v projektu in po koncu projekta.

Natančneje:

- Kako bomo opisali, analizirali in hranili podatke v projektu?
- Na kak način bomo arhivirali in delili podatke izven raziskovalne skupine?





## Načrt ravnanja s podatki

NRRP obrazec lahko **predpiše financer ali raziskovalna organizacija**. Zato velja pred sprejemom odločitve glede izbora obrazca **preveriti zahteve, določene v pogodbi o financiranju** raziskave in/ali pravila raziskovalne organizacije.

- NRRP je "**živ**" dokument, ki se spreminja skupaj s potrebami projekta in njegovih udeležencev.
- Med projektom se NRRP **posodablja**, da se zagotovi sledenje spremembam skozi čas in da odraža trenutno stanje raziskovalnega projekta.
- Izpolnjeni načrti ravnanja z raziskovalnimi podatki se med seboj **zelo razlikujejo**, saj so vedno oblikovani na podlagi posebnih potreb podatkov, zbranih v okviru vašega projekta.



# Obdelava in analiza podatkov

- Vnos podatkov, digitalizacija, prepis, prevod.
- Preverjanje, potrjevanje, čiščenje, anonimizacija podatkov, kjer je potrebno.
- Opis, strukturiranje in dokumentiranje podatkov.
- Analiza podatkov.
- Interpretacija rezultatov.
- Upravljanje s podatki in hramba (varnostne kopije: 3-2-1 pravilo arhiviranja).





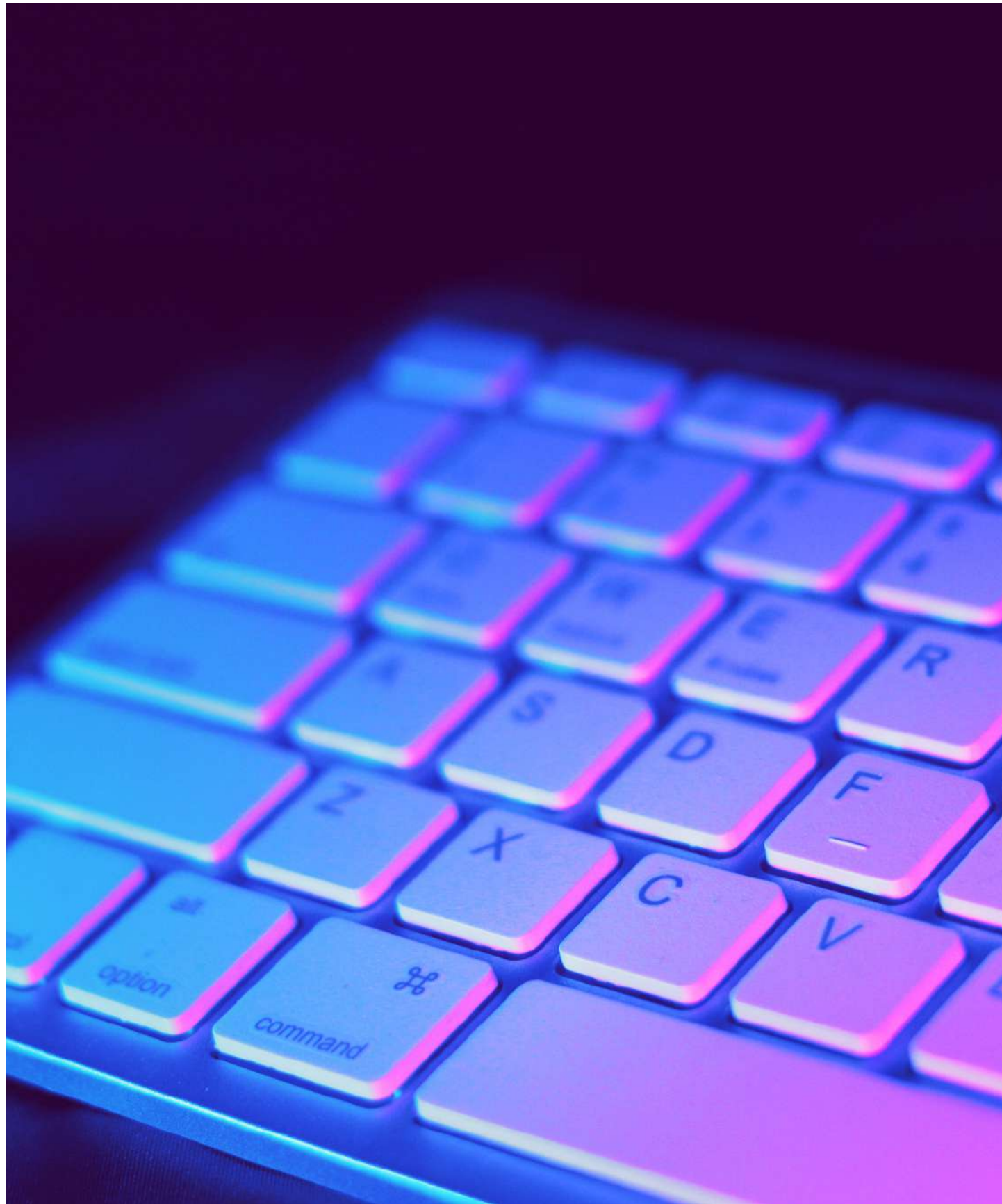


# Objava podatkov

Podatkovno objavo je treba obravnavati kot prvovrsten raziskovalni rezultat (Knowledge Exchange, 2013).

Da bi zbirka podatkov bila objava, mora biti:

- ustrezno dokumentirana z metapodatki,
- pregledana glede kakovosti,
- mogoče jo je poiskati in najti v katalogih (ali podatkovnih zbirkah),
- opremljena z informacijami za citiranje v člankih.



# Vrste dostopa

- Odprti dostop
- Standardni dostop ob registraciji
- Dostop pod posebnimi pogoji
  - a) datoteka za raziskovalne namene (SUF)
  - b) datoteka v varovanem okolju (SCUF)

## **Pomembno:**

**Preverim, ali mi financer/delodajalec/fakulteta predpisuje**

- **točno določen obrazec NRRP**
- **obdobja posodabljanja obrazca NRRP**
- **rok za oddajo obrazca NRRP**

## **Pomembno:**

**Preverim moje obveznosti glede ravnanja z raziskovalnimi podatki**

- **Ob podpisu pogodbe o financiranju raziskave**
- **Ob zaposlitvi na raziskovalni organizaciji**
- **Ob vpisu na doktorski študij**
- **Navodila za avtorje znanstvenih člankov pri znanstvenih revijah**





**Vprašanja?**

**HVALA**