

# Pētniecības datu pārvaldība pa soļiem

Ingmārs Kreišmanis,  
Zinātnes departaments  
Zinātnieku brokastis  
2022. gada 2. novembris

RSU

## PAR KO RUNĀSIM?

### Kāpēc pētniecības datu pārvaldība?

- Atvērtā zinātne un atvērtie dati
- Atbildīga pētniecība un inovācijas
- FAIR principi

### Pētniecības datu dzīves cikls

- Plānošana
- Datu ievākšana, apstrāde, uzglabāšana
- Datu publicēšana un atkārtota izmantošana

### Aktualitātes Latvijā un RSU

- Latvijas atvērtās zinātnes stratēģija (2021.-2027.)
- Aktuālie darbi RSU ietvaros
- Atbalsts un resursi

**PĒTNIECĪBAS DATU  
PĀRVALDĪBA PA  
SOĻIEM**

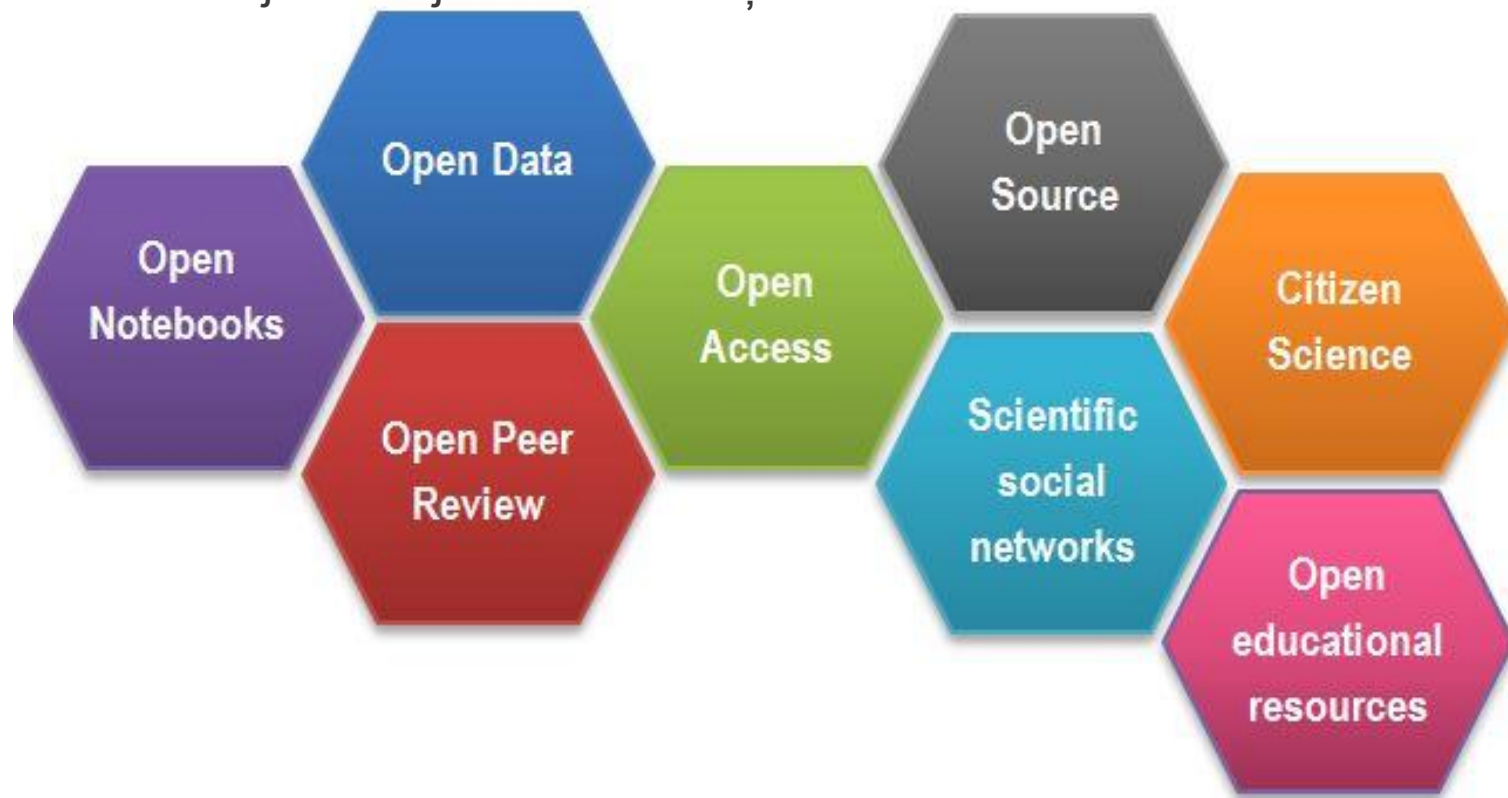
# **Kāpēc pētniecības datu pārvaldība?**



**RĪGAS STRADIŅA  
UNIVERSITĀTE**

# Atvērtā zinātne un atvērtie dati

«Dati ir atvērti, ja tiem var brīvi piekļūt, izmantot, modificēt un dalīties ar jebkuru jebkādam mērķim»



DRS

# Atbildīga pētniecība un inovācijas

## Galvenie principi:

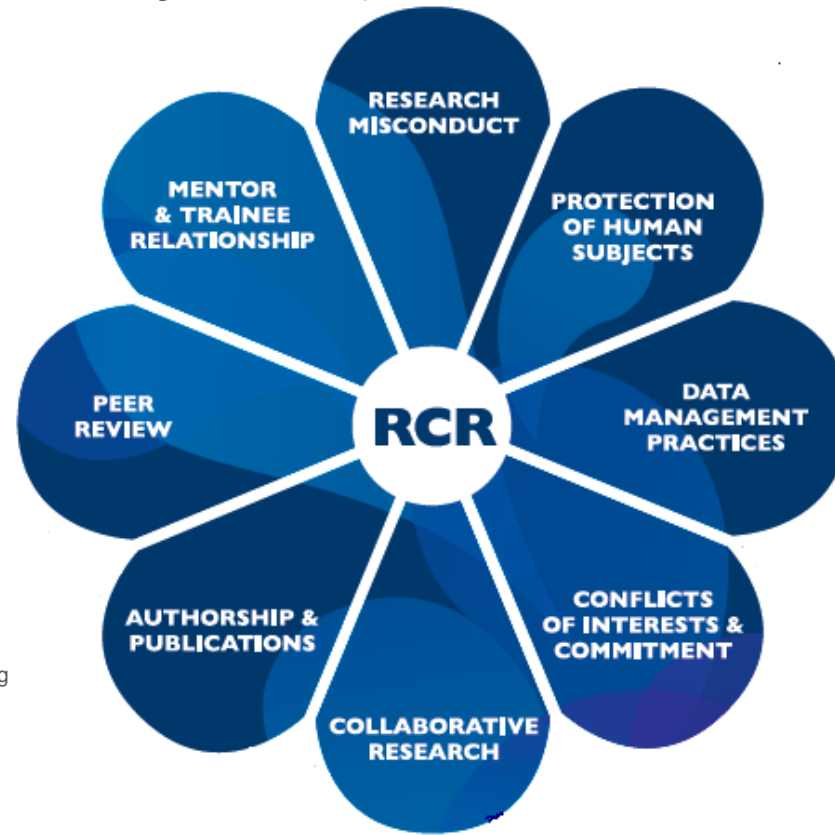
- pētniecībai vajadzētu sniegt labumu sabiedrībai
- pētniecībai nevajadzētu radīt kaitējumu
- cieņa – pētniecībai un pētniecības videi vajadzētu atbalstīt cieņu pret indivīdiem
- taisnīgums – pētniecībai vajadzētu palīdzēt pasaulei kļūst taisnīgākai

## Galvenās vērtības:

- atvērtība
- caurspīdīgums
- reproducējamība

Avots: prezentācija «Open Science, Open Access, Open Data», Louise Bezuidenhout, Data Archiving and Networked Services

Atbildīga pētniecības veikšana (*responsible conduct of research*)



# FAIR principi

## Findable:

- F1.** (meta)data are assigned a globally unique and persistent identifier;
- F2.** data are described with rich metadata;
- F3.** metadata clearly and explicitly include the identifier of the data it describes;
- F4.** (meta)data are registered or indexed in a searchable resource;

## Interoperable:

- I1.** (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2.** (meta)data use vocabularies that follow FAIR principles;
- I3.** (meta)data include qualified references to other (meta)data;

## Accessible:

- A1.** (meta)data are retrievable by their identifier using a standardized communications protocol;
  - A1.1** the protocol is open, free, and universally implementable;
  - A1.2.** the protocol allows for an authentication and authorization procedure, where necessary;
- A2.** metadata are accessible, even when the data are no longer available;

## Reusable:

- R1.** meta(data) are richly described with a plurality of accurate and relevant attributes;
  - R1.1.** (meta)data are released with a clear and accessible data usage license;
  - R1.2.** (meta)data are associated with detailed provenance;
  - R1.3.** (meta)data meet domain-relevant community standards;

## FAIR principi:

Atrodami  
(*Findable*)

Pieejami  
(*Accessible*)

Sadarbspējīgi  
(*Interoperable*)

Atkārtoti  
izmantojami (*Re-usable*)

doi: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)

Slide CC-BY by Erik Schultes, Leiden UMC

**PĒTNIECĪBAS DATU  
PĀRVALDĪBA PA  
SOĻIEM**

# **Pētniecības datu dzīves cikls**

# Galvenie soļi pētniecības datu pārvaldībā

## Datu pārvaldības plāns, plānošana, ētiskie un juridiskie aspekti

- resursu apzināšana datiem, institūcijas iekšējie nosacījumi, DMP
- atļaujas no ētikas komisijām
- informētā piekrišana

## Datu sagatavošana

- programmatūras izvēle (ieteicams atvērtā koda un brīvpieejas/nozarē plaši izmantota)
- datu tīrīšana un apstrāde, izmantojot nozarei un tematikai atbilstošās konvencijas
- personas datu un sensitīvas informācijas gadījumā pseidonimizācija vai anonimizācija

## Failu formāti un failu organizēšana

- datu saglabāšana brīvi pieejamā un nozarei izplatītā datnes formātā
- failu organizēšana pēc vienota principa (nodrošinot izsekojamību arī sev), versiju kontrole

## Metadati un datu dokumentēšana

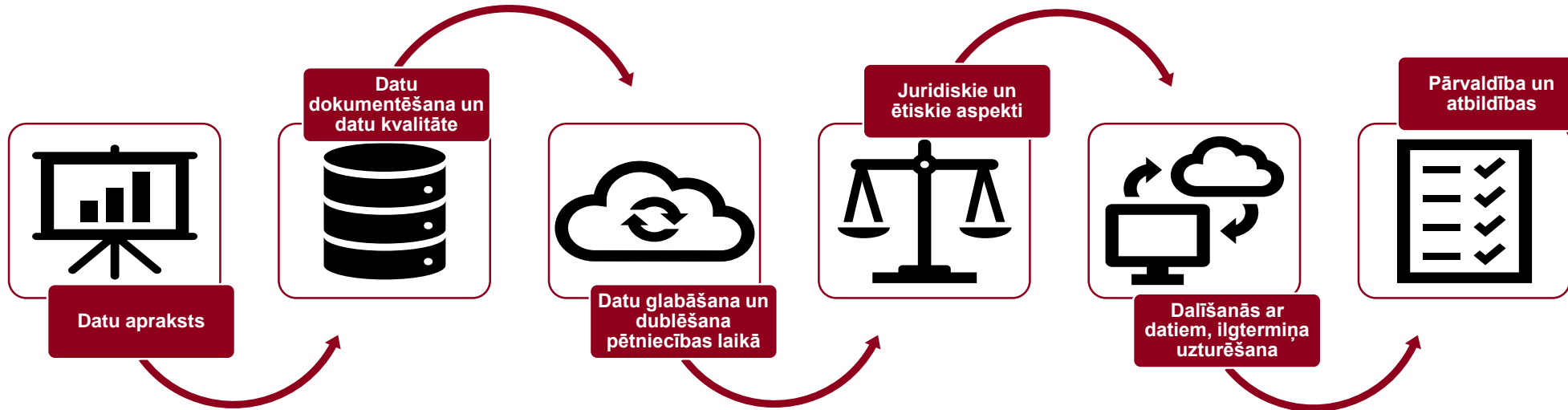
- metadatu apraksts atbilstoši nozares standarta klasifikācijām (ja repozitorijs nepieļauj visus vajadzīgos laukus, var pievienot arī .txt dokumentu)
- nodrošināt dokumentāciju par datiem (*codebook*, *ReadMe* fails, metodoloģijas apraksts, anketas jautājumi, programmatūras kods u.c.)

## Datu publicēšana un uzturēšana

- datu publicēšana (atvērtā vai ierobežotā piekļuvē) atbilstošā repozitorijā (ieteicams, ka automātiski piešķir PID (piem., DOI), licenci (piemēram, *Common Creatives*), ir atvērts un pieejams protokols datu un metadatu pieejamībai)
- jānosaka laiks datu uzturēšanai (atkarīgs no institūcijas un repozitorija politikas)
- nepieciešamības gadījumā labojumi un papildinājumi (nodrošinot skaidras atzīmes par versijām un veiktajām izmaiņām)
- atsauces uz datiem zinātniskajos rakstos



# Datu pārvaldības plāns (DMP)



# Datu sagatavošana

Visā datu dzīves ciklā jāatceras ielānot gan finanšu, gan cilvēkresursus, to attiecīgi ielānojot arī projektā. Jāatceras, ka arī pēc projekta beigām var būt izmaksas.

Lai nodrošinātu pienācīgu datu kvalitāti, jāsak plānot jau pirms datu ievākšanas/generēšanas, nosakot pamatprincipus datu strukturēšanai, validācijai un aprakstam

Jāizvēlas atbilstoša programmatūra datu ievadīšanai apstrādes vidē (atkarībā no nozares tā var būt laboratorijas programmatūra, aptauju rīki, transkripcijas rīki u.c.)

Personas dati un sensitīvi dati jāaizsargā visā pētniecības laikā, tos apstrādājot drošā vidē, kā arī pseidonimizējot tos pēc iespējas ātrāk

Datus jāapraksta un jākodē atbilstoši nozares tradīcijām (tajā skaitā kodēt necipariskās vērtības atbilstoši standartiem)

Visiem šiem soļiem iespējams piemēklēt atbilstošu programmatūru, piemēram, *RedCap*, *LIMS*, *SPSS* u.c. atbilstoši konkrētajām vajadzībām

# Failu formāti un organizēšana

**Mērķis padarīt datus viegli pieejamus un iegūstamus arī ilgtermiņā, tādēļ:**

- jāizvēlas datnes formāti, kas ir nesaspiesti, atverami ar nekomerciālām programmām, atbilstoši starptautiskiem standartiem
- jāpiemēro konsekventa pieeja savu failu organizēšanai, vienotu pieeju failu nosaukšanai
- failu atbilstoša organizēšana un piemēroti datņu formāti palīdzēs arī pašiem pētniekiem orientēties un iegūt/apskatīt datus pēc ilgāka laika posma

## Biežākie gadījumi

**Piemēroti datnes formāti**

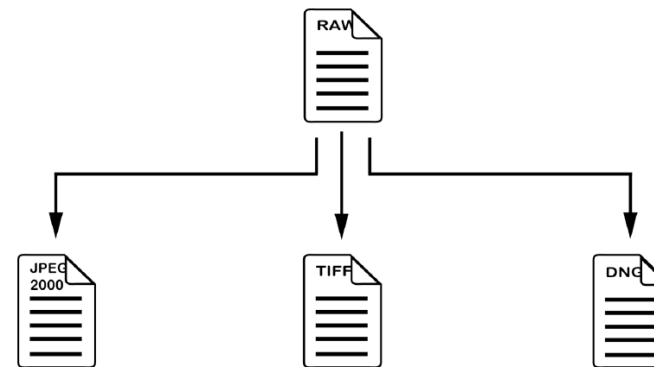
CSV, TAB

PDF, TXT

**Nepiemēroti datnes formāti**

XLSX

WORD



cc by 4.0 Raman Ganguly

# Metadati

**Metadatiem («dati par datiem») ir vairāki uzdevumi:**

- aprakstīt digitālos objektus (piemēram, datu kopas)
- nodrošināt šo objektu pieejamību un juridisko drošību
- iespējot tālāku objektu izmantošanu un tie ir brīvi pieejami/izmantojami
- nodrošināt apmaiņu ar citām sistēmām

Avots: prezentācija «Introduction to Research Data Management», Susanne Blumesberger (Vienna University, Vienna University Library)



Avots: <https://dmeg.cessda.eu/Data-Management-Expert-Guide/2.-Organise-Document/Documentation-and-metadata>

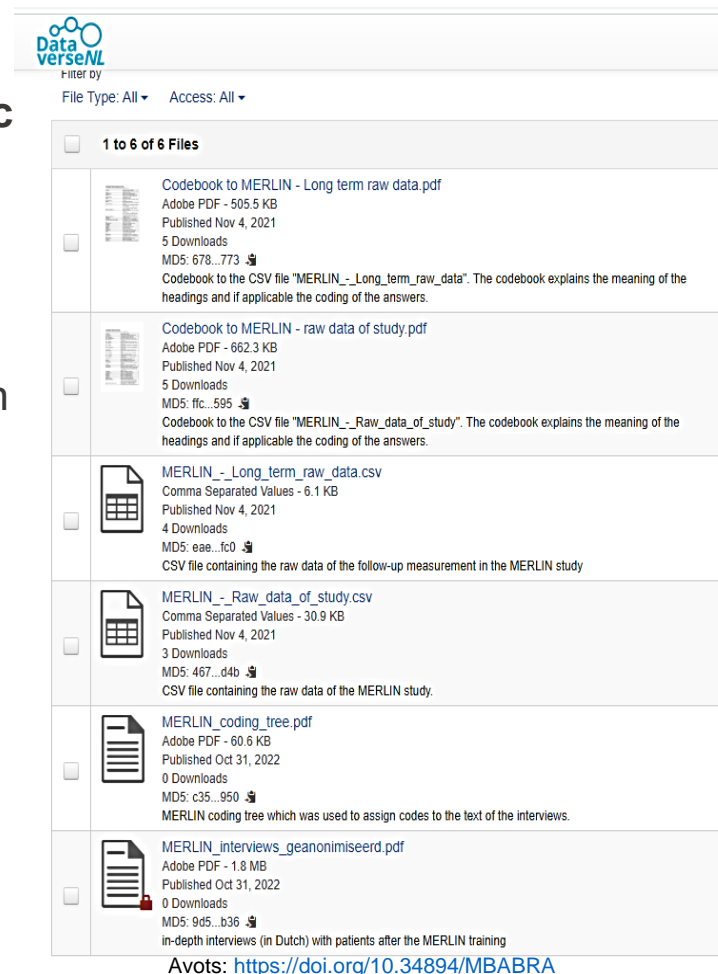
**Populārākās metadatu shēmas** Dublin Core, Data Documentation Initiative (DDI) u.c.

Metadatus var pievienot arī failu veidā (piemēram, .txt datnes formātā), ja attiecīgajā repozitorijā nav iespējams ievadīt visus nozarei raksturīgos metadatus, sevišķi ja ir pieejami specifiski standarti. Dažādus standartus var atrast [šeit](#) un citās vietnēs.

# Datu dokumentēšana

Pētniecības datus ir jādokumentē, lai tie būtu pēc iespējas reproducējami nākotnē, piemēram:

- metodoloģijas apraksti – pētījuma un datu vākšanas vai ģenerēšanas metodoloģija
- kodu grāmatas – datu tehniskais apraksts, izmantotie mainīgie un skaitļi, datu kopu struktūra un cita konteksta informācija
- anketas – sevišķi aptaujas datu gadījumā ir svarīgi pievienot anketu datnes
- laboratorijas piezīmju grāmatas un eksperimentālie protokoli – piezīmes, kurās dokumentēti pētījumi
- ar programmatūru saistīta dokumentācija – ieteicams pievienot dokumentāciju un kodu
- readme.txt – datne ar norādījumiem, kā reproducēt jūsu pielietoto datu analīzi



The screenshot displays the DataVerseNL interface for a file repository. At the top, there is a logo for DataVerseNL and a filter section with 'File Type: All' and 'Access: All'. Below this, a header indicates '1 to 6 of 6 Files'. The main content area lists six files, each with a checkbox, a file icon, a title, file type, size, publication date, download count, MD5 hash, and a brief description.

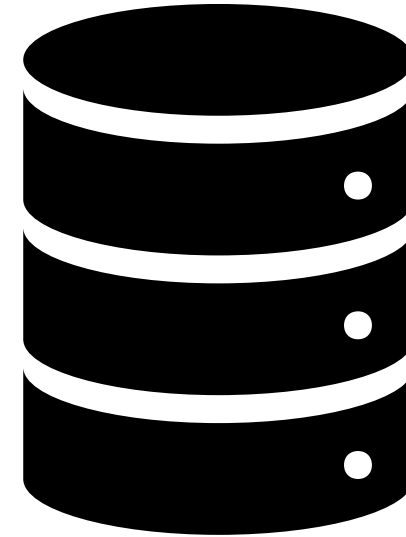
File Name	Type	Size	Published	Downloads	MD5	Description
Codebook to MERLIN - Long term raw data.pdf	Adobe PDF	505.5 KB	Nov 4, 2021	5	678...773	Codebook to the CSV file "MERLIN_-_Long_term_raw_data". The codebook explains the meaning of the headings and if applicable the coding of the answers.
Codebook to MERLIN - raw data of study.pdf	Adobe PDF	662.3 KB	Nov 4, 2021	5	ffc...595	Codebook to the CSV file "MERLIN_-_Raw_data_of_study". The codebook explains the meaning of the headings and if applicable the coding of the answers.
MERLIN_-_Long_term_raw_data.csv	Comma Separated Values	6.1 KB	Nov 4, 2021	4	eae...fc0	CSV file containing the raw data of the follow-up measurement in the MERLIN study
MERLIN_-_Raw_data_of_study.csv	Comma Separated Values	30.9 KB	Nov 4, 2021	3	467...d4b	CSV file containing the raw data of the MERLIN study.
MERLIN_coding_tree.pdf	Adobe PDF	60.6 KB	Oct 31, 2022	0	c35...950	MERLIN coding tree which was used to assign codes to the text of the interviews.
MERLIN_interviews_geanonimiseerd.pdf	Adobe PDF	1.8 MB	Oct 31, 2022	0	9d5...b36	In-depth interviews (in Dutch) with patients after the MERLIN training

Avots: <https://doi.org/10.34894/MBABRA>

# Datu publicēšana un uzturēšana

## Prasības pētniecības datu repozitorijam:

- nodrošina drošu glabāšanu, veicot regulāras datu dublēšanas
- piedāvā atvērtus standartus piekļuvei datiem un metadatu standartus
- piešķir pastāvīgo unikālo identifikatoru (piemēram, *DOI*)
- ir reģistrēts vai sertificēts (piemēram, *Re3Data*, *EOSC* portālā, *CoreTrustSeal*)
- piešķir licences (piemēram, *Common Creatives*)
- ir skaidrība par datu un metadatu uzglabāšanas pieeju un ilgumu
- ja attiecināms, nodrošina iespējas uzglabāt sensitīvu informāciju ierobežotā vai slēgtā piekļuvē
- labojumu gadījumā nodrošina automatizētu versiju noteikšanu un aprakstu



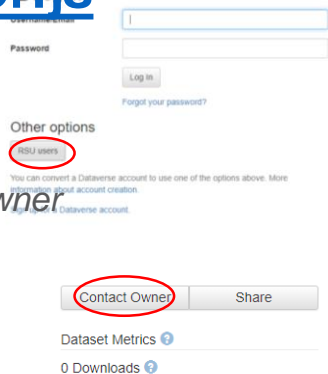
**RSU ietvaros datu saglabāšana notiek institucionālajā pētniecības datu repozitorijā Dataverse!!!**

# RSU Dataverse

## RSU institucionālais pētniecības datu repozitorijs

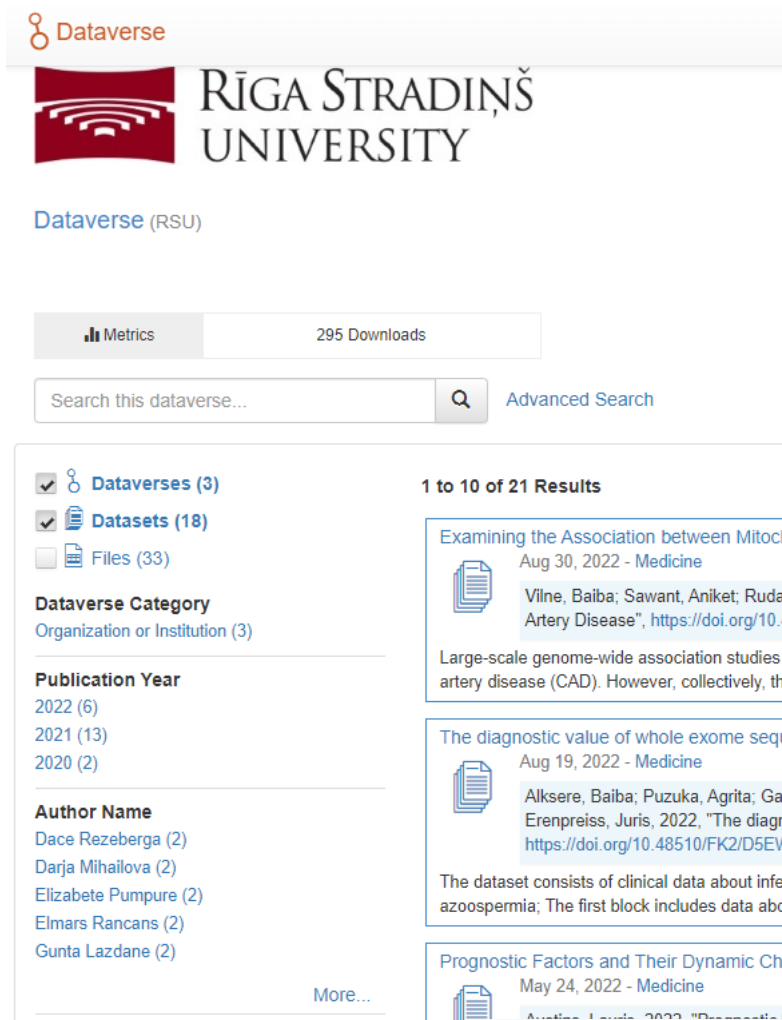
Meklēt un izmantot datu kopas:

- piekļuve ar RSU autentifikāciju
- RSU pētnieku izstrādātās datu kopas
- ja ierobežota piekļuve, spiest *Contact owner*
- tiks regulāri papildināts ar kopām no aktuālajiem projektiem



Publicē un popularizē savas datu kopas:

- atvērti metadatu standarti (DDI), dublēšana, licence (CC0), unikālais identifikators, var nodrošināt ierobežotu piekļuvi, automatizēta versiju veidošana u.c.
- par datu kopas ievietošanu rakstīt [dataverse@rsu.lv](mailto:dataverse@rsu.lv)
- iespējas popularizēt, Dataverse reģistrēts nozīmīgākajās resursu datubāzēs



# Pētniecības datu pieejamība

**«Tik atvērti, cik iespējams, tik slēgti, cik nepieciešams»:**

- atvērta piekļuve (visiem ir piekļuve, datiem nevajadzētu saturēt personas datus)
- ierobežota/slēgta piekļuve (piekļuve tikai pēc pieprasījuma un apstiprinājuma, metadati pieejami)
- embargo periods – iespējams datus atklāt pēc noteikta laika (piemēram, kad publicēti citi zinātniskie rezultāti, metadati pieejami)

**Uz publicētajām datu kopām ieteicams atsaukties citos zinātniskajos rezultātos, tas palīdzēs:**

- popularizēt datus
- veikt zinātnisko rakstu un to satura validāciju
- veicināt jaunas sadarbības iespējas





# Datu atkārtota izmantošana

Ir ieteicams atkārtoti izmantot jau izstrādātas datu kopas, jo:

- samazinās pētnieku darbu un universitātes resursu izmantošanu
- būs vairāk laika analīzei
- var veicināt sadarbības iespējas, ja izmantojat citu pētnieku datu kopas
- studējošie var izmantot jau izstrādātas kvalitatīvas datu kopas



Lai atkārtoti izmantotu kādu datu kopu, jāizvērtē:

- piekļuves un izmantošanas tiesības (licences veids);
- datu kvalitāte;
- dokumentācijas kvalitāte;
- resursa uzticamība



**PĒTNIECĪBAS DATU  
PĀRVALDĪBA PA  
SOĻIEM**

# **Aktualitātes Latvijā un RSU**



**RĪGAS STRADIŅA  
UNIVERSITĀTE**

# Latvijas atvērtās zinātnes stratēģija (2021.-2027.)

**Ieviesti standartizēti datu pārvaldības plāni pētījumu projektu konkursos**

*Principi: minimāls slogs pētniekiem, starptautiski šabloni, cik vien iespējams automatizēti*

**DataverseLV pētniecības datu repozitoriju tīkls**

Federēta sistēma, kurā tiks glabāti arī RSU Dataverse datu kopu metadati

**Dalība Eiropas Atvērtās zinātnes mākonī (EOSC)**

**Izveidots Augstākās izglītības un zinātnes koplietošanas IT pakalpojumu centrs (VPC)**

The logo for 'argos' is written in a lowercase, rounded, teal-colored font.The logo for 'DataverseLV' features the word 'Dataverse' in a grey sans-serif font, followed by 'LV' in a larger, bold, red font. The 'V' has three small red circles above it, resembling a molecular structure or data points.

**EUROPEAN OPEN  
SCIENCE CLOUD**

# Pētniecības datu pārvaldība RSU – aktuālie darbi

Datu kopu identificēšana un publicēšana RSU – publicēšana [Dataverse](#) un [PURE](#), vienota metadatu anketa



Atbalsta un apmācību plāns RSU pētniekiem, lai uzlabotu pētniecības datu kvalitāti un pārvaldības pasākumus

Iekšējie normatīvie akti un metodoloģiskie materiāli pētniekiem



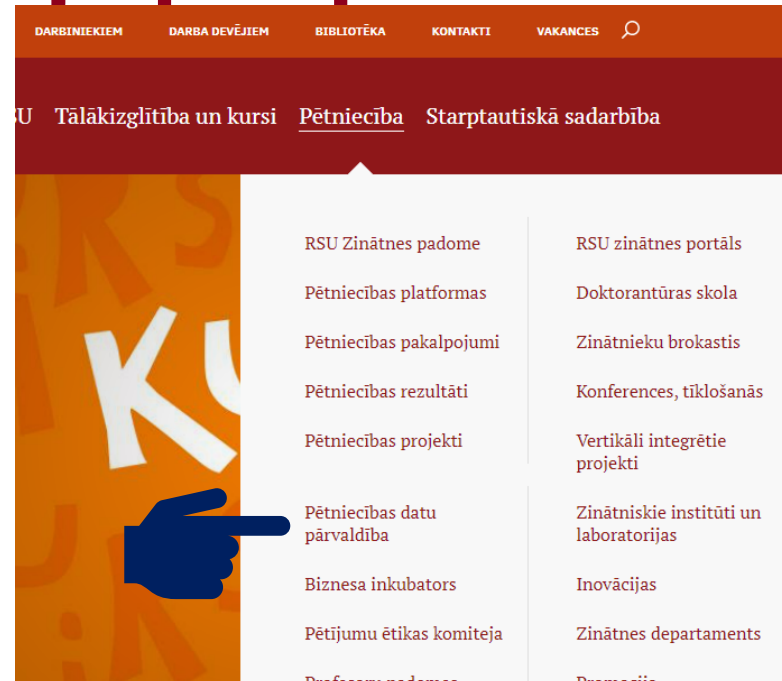
Sadarbība ar [European Open Science Cloud](#) (EOSC), RSU dalība un projekti

RSU datu pārvaldības plāna forma (DMP), [Argos](#) izmantošana DMP izveidei un iesniegšanai



# RSU mājaslapas sadaļa par pētniecības datu pārvaldību

- Pirmie soļi pētniecības datu pārvaldībā
- FAIR principi
- Datu pārvaldības plāns
- Ieteikumi un saites uz noderīgiem resursiem
- Ko papildināt? – droši rakstiet [ingmars.kreismanis@rsu.lv](mailto:ingmars.kreismanis@rsu.lv)
- [LV versija](#) un [ENG versija](#)



# DMP – RSU veidne

## RSU DMP veidnē iekļautas šādas sadaļas:

- datu kopsavilkums
- FAIR dati
- resursi un drošība

Datu pārvaldības plāns palīdzēs pētniekiem plānot visas darbības attiecībā uz datu vākšanu (tostarp ētiskiem un tehniskiem nosacījumiem), datu apstrādi un publicēšanu



argos

✓ General Info

### 1 Data Management Plan (...)

1.1 Data Summary 8

### 2 FAIR data

2.1 Making data finda... 7

2.2 Making data openl... 7

2.3 Making data intero... 4

2.4 Increase data re-use 6

### 3 Resources and Security

3.1 Allocation of resou... 4

3.2 Data security 3

3.3 Ethical aspects 6

DRS

# Pasākumi novembrī



## FAIR datu principu piemērošana sociālajās zinātnēs

Laiks: 14. un 23. novembris  
Vieta: tiešsaistē, būs pieejams ieraksts e-studijās

Uzstāsies:  
Lubļanas universitātes docents un Slovēnijas sociālo zinātņu datu arhīva vadītājs **Janez Štebe**

Tematika: kā dokumentēt un aprakstīt savus datus, lai tie būtu izmantojami un sadarbspējīgi

Vairāk informācijas [šeit](#)

## FAIR datu principu piemērošana medicīnas un veselības zinātnēs

Laiks: 15. un 22. novembris  
Vieta: tiešsaistē, būs pieejams ieraksts e-studijās

Uzstāsies:  
**Gerijs Sanderss** (Gary Saunders), EATRIS (European infrastructure for Translational medicine) datu direktors  
**Ronalds Kornets** (Ronald Cornet), Amsterdamas Universitātes Sabiedrības veselības pētniecības institūta asociētais profesors

Tematika: kā dokumentēt un aprakstīt savus datus, lai tie būtu izmantojami un sadarbspējīgi

Vairāk informācijas [šeit](#)

**Paldies par uzmanību!**

Jautājumiem un konsultācijām:

[ingmars.kreismanis@rsu.lv](mailto:ingmars.kreismanis@rsu.lv)

RSU