

No datu pārvaldības pamatiem līdz datu publicēšanai

Monta Simona Zeltiņa, RSU datu kuratore

Latvija datu kuratoru tīkls, CC BY 4.0

Vebinārs «Pētniecības datu ekosistēma Latvijā un bibliotēku loma» 22.04.2026.



Projekts «Atbalsts atvērtās zinātnes ieviešanai praksē, kā arī izveidoti risinājumi zinātnes datu koplietošanai un dalībai ES atvērtajā zinātnes mākonī» (ANM 2.1.3.1.i.)

Pētniecības datu pārvaldība un FAIR principi



Pētniecības dati

Piemēri:

- eksperimentu dati
- aptauju rezultāti
- intervijas
- novērojumi
- teksti
- attēli
- programmas kods
- klīniskie pētījumi
- laika apstākļu dati
- astronomiskie dati
- simulāciju dati
- u.c.

Pētniecības dati ir jebkāda informācija, kas tiek savākta, apstrādāta un analizēta, lai atbildētu uz pētniecības jautājumiem.

- Jēldati jeb neapstrādāti dati (raw data)
- Tīrīti dati
- Apstrādāti dati

- Atvērtie dati
- Kopīgoti dati
- Slēgtie dati
- Dati, kuriem noteikts embargo periods

Datu kopa ir sistemātiski strukturēts datu kopums un sastāv no vairākiem datu elementiem vai vērtībām.

- Primārie dati
- Sekundārie dati

- Kvantitatīvie dati
- Kvalitatīvie dati

- Dati: "25", "Latvija", "Sieviete"
- Datu kopa: tabula ar datiem



Sensitīvi dati

Informācija, kuras neatbilstoša izmantošana var radīt kaitējumu personai, tās tiesībām, brīvībām, reputācijai vai drošībai.

Sensitīvo datu veidi

- Personas dati
- Specifiski vides dati
- Duālā pielietojuma dati
- Komerccnoslēpumi

Normatīvais regulējums

- VDAR un Fizisko personas datu apstrādes likums
- Institūciju ētikas un drošības vadlīnijas
- Finansētāju prasības
- Specifiskie tiesību akti

Sensitīvo datu formāti

- Skaitliskie dati
- Tekstuālie dati
- Attēli
- Audio
- Citi formāti

Pētniecības datu pārvaldība

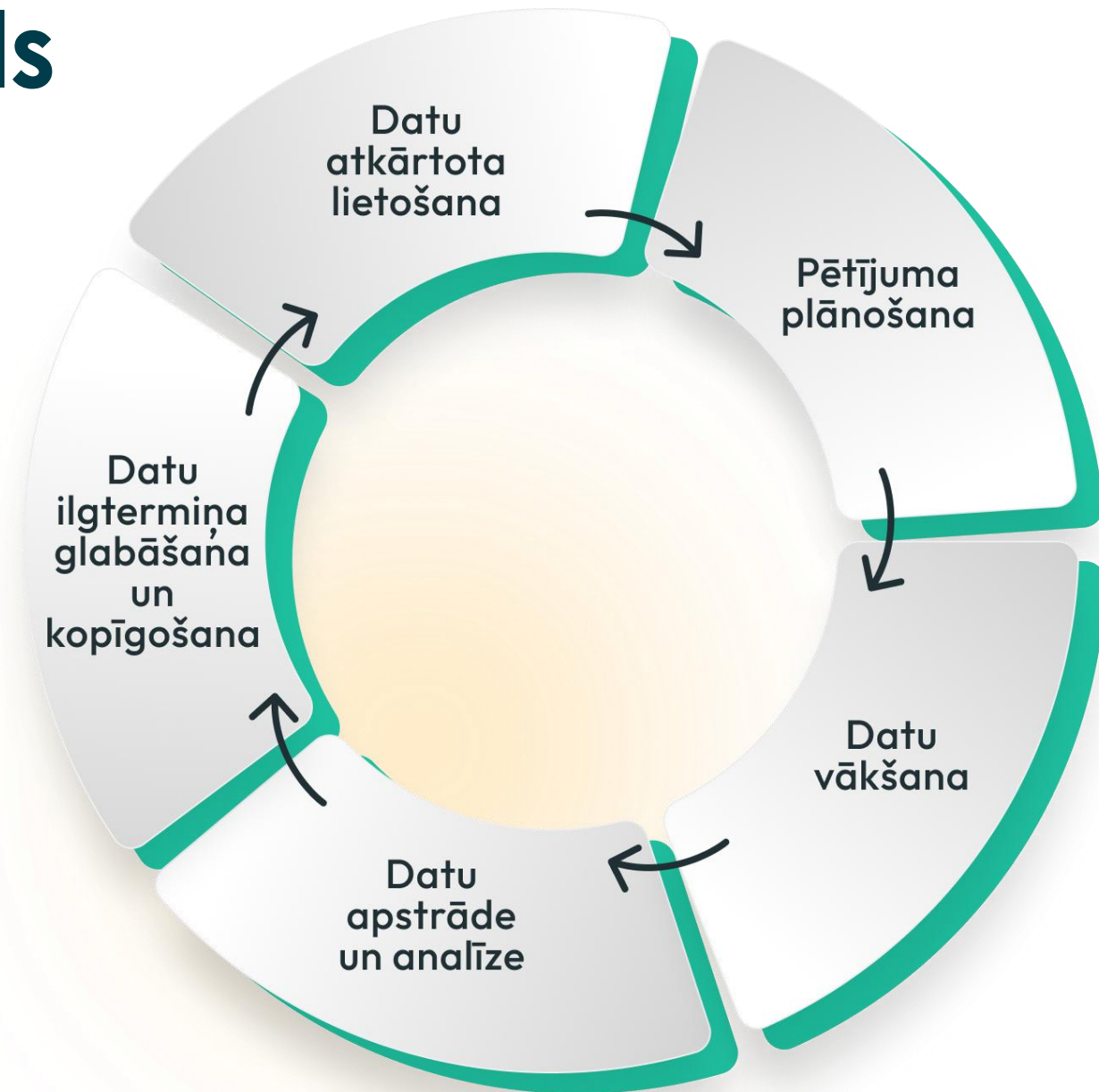
Pētniecības datu pārvaldība ir process, kurā sistemātiski tiek plānota, organizēta, uzglabāta, koplietota un saglabāta pētniecības laikā iegūtā informācija jeb dati.

- ✓ **Tehniskie, organizatoriskie, strukturālie, likumdošanas un ilgtspējības** aspekti
- ✓ Laikus saplānots pētījums – **ietaupījums ilgtermiņā**
- ✓ **Datu pārvaldības plāns** – visa informācija par datu pārvaldību vienuviet



Datu dzīves cikls

- ✓ Datu integritāte
- ✓ Reproducējamība
- ✓ Prasību ievērošana
- ✓ Efektivitāte
- ✓ Sadarbība



FAIR principi

F

Findable

Dati un to metadati ir viegli **atrodami** citiem pētniekiem un sistēmām

A

Accesible

Dati ir **pieejami** un piekļuves nosacījumi ir skaidri norādīti

I

Interoperable

Dati ir **savietojami** ar citām sistēmām un datu kopām

R

Reusable

Dati ir sagatavoti tā, lai tos varētu **atkārtoti izmantot** nākotnē

FAIR dati ≠ Atvērti dati

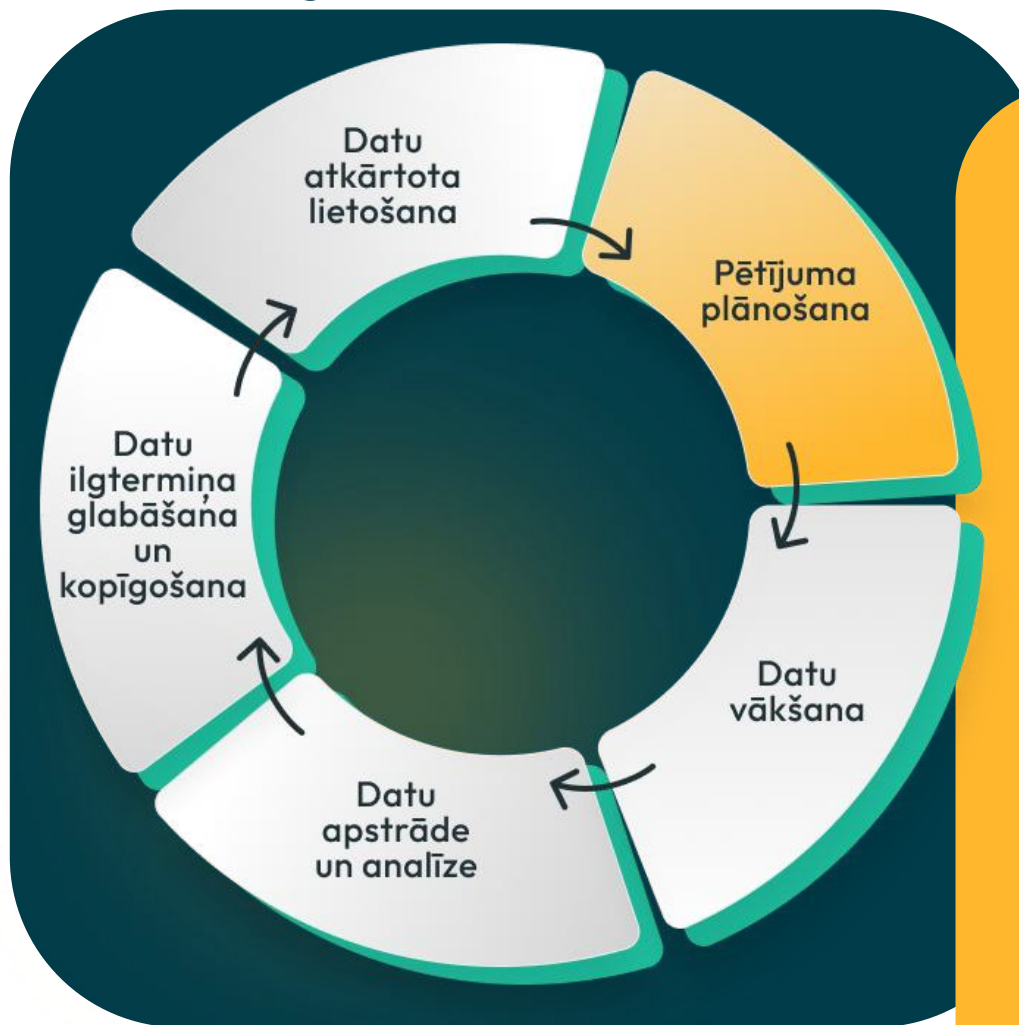
Atvērti dati ≠ Kopīgoti dati

*Tik atvērti, cik iespējams,
tik slēgti, cik nepieciešams.*



Pētniecības datu dzīves cikls un datu pārvaldība ikdienā

Pētījuma plānošana



Datu pārvaldības plāns – datu vākšana, organizēšana, aprakstīšana, piekļuve, droša īstermiņa un ilgtermiņa glabāšana, publicēšanas kārtība, nepieciešamie resursi, izmaksas

Ētikas komitejas atzinums, citas formālās prasības – pētījums atbilst normatīviem, ievērots dalībnieku privātums un brīvprātīga dalība

Datu glabāšana, apstrāde un piekļuves kontrole – kur dati tiks glabāti, kā apstrādāti, kurš varēs tiem piekļūt

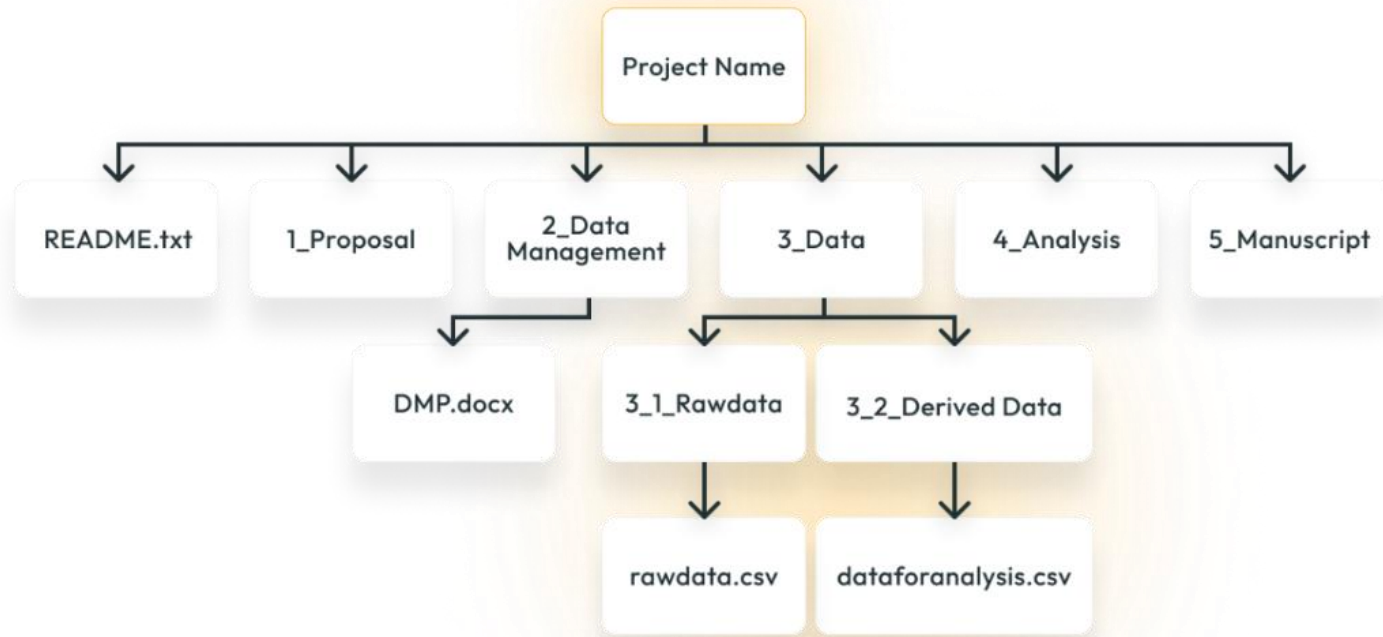
Datu pārvaldības plāns

- ✓ Vispārīga **informācija par projektu**, tā mērķi un finansējošo organizāciju
- ✓ **Kādi dati** tiks ģenerēti un izmantoti?
- ✓ Kā dati tiks **organizēti**?
- ✓ **Kā dati tiks aprakstīti**, ar kādiem metadatiem un dokumentāciju?
- ✓ Kāda ir **datu glabāšanas un drošības** stratēģija un risinājumi?
- ✓ Kā, kad un **kur dati tiks publicēti**, ar kādiem nosacījumiem ir paredzēts dalīties ar tiem?
- ✓ Datu pārvaldībai nepieciešamās **izmaksas un resursi**
- ✓ **Ētiskas un intelektuālā īpašuma** jautājumi

DPP ir "dzīvs" dokuments – tas var tikt labots un papildināts visā projekta gaitā.



Datņu organizēšana un nosaukšana



Ieteicams izvairīties no īpašām rakstzīmēm, simboliem un diaktriskajām zīmēm: **?: * "> < | : #%" } | ^ [] ~ œÆ øØ äÅ ņÑ šŠ**

Jāpievieno datņu versiju apzīmējumus: **v001, v002, v003**

Datumus jānorāda pēc starptautiskas konvencijas: **2017-10-25**

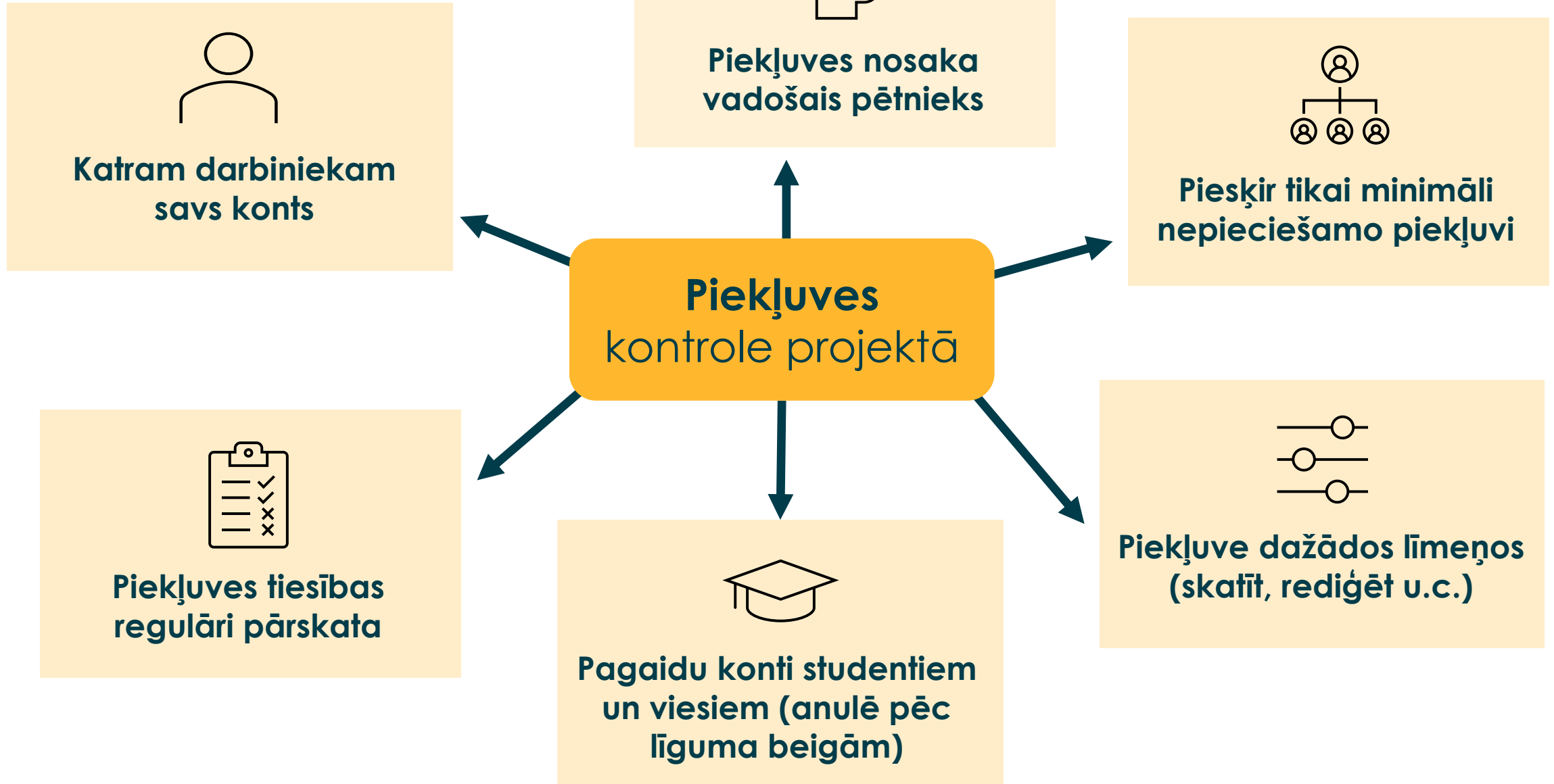
Atstarpes jāaizstāj: **kodu_gramata, kodu-gramata, KoduGramata**

Labā prakse:

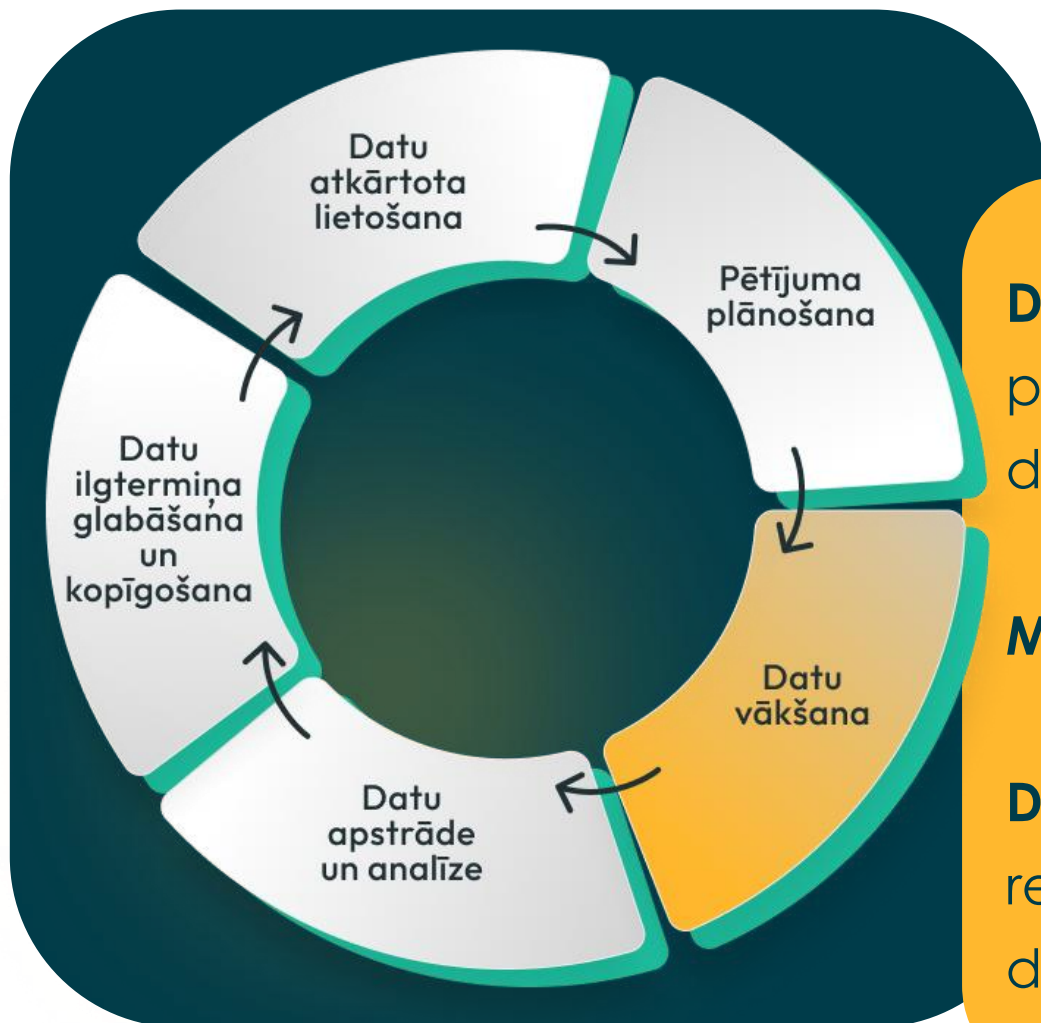
- apakšmapes ne vairāk kā 5 līmeņos
- jāizvairās no liekas informācijas dublēšanās datņu nosaukumos un struktūrā
- loģiska secība un numerācija:
 - 01_admin
 - 02_dati
 - 03_datu_analize
 - 04_rezultati
- atsevišķa mape (vairs nevajadzīgo) datņu arhivēšanai

Jāievēro konsekventa pieeju un jādeleģē atbildīgā persona!

Piekļuves kontrole



Datu vākšana



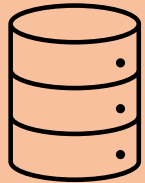
Droša un ētiska datu ievākšana – dalībnieku informētā piekrišana, datu avotu prasības (piem., uzņēmumu datiem)

Metožu un rīku izvēle – droši un validēti rīki

Datu uzglabāšana – drošas glabātuves (institūcijas resursi), datu aprakstīšana (metadati un dokumentācija)

Datu īstermiņa glabāšana

Pētniecības **dati**
iestādes infrastruktūrā



Privātie datori
un serveri



USB zibatmiņas



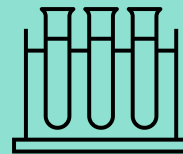
Google Drive



Kopīgoti e-pastā vai privātos
mākoņservisos



Iestādes serveri



Iekārtas savienotas ar serveri



SharePoint



Google Drive

Iestādes mākoņpakalpojumi
(ar iestādes lietotājvārdu)

Aizsargājamiem datiem īpaša
vieta un **ierobežota** piekļuve

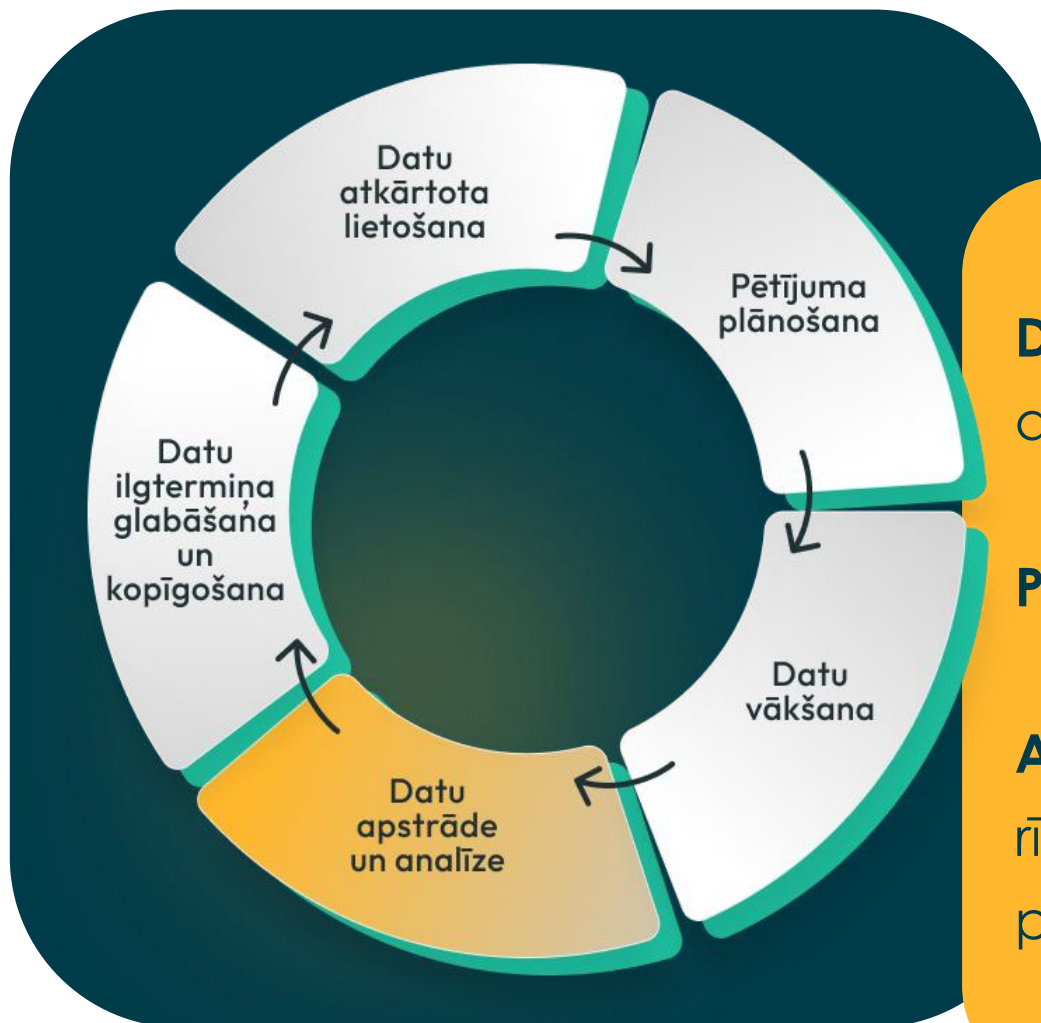


leguvumi



- Dati pasargāti no zaudēšanas
- Uzlabota datu drošība
- Labāk nodrošināta piekļuves kontrole
- Vieglāk izpildīt finansētāju un normatīvās prasības

Datu apstrāde un analīze



Dokumentēšana – rūpīgi dokumentēti apstrādes un analīzes soļi

Pseudonimizācija & anonimizācija – sensitīviem datiem

Apstrādei un analīzei – vēlams izmantot brīvpieejas rīkus, lai to varētu atkārtot jebkurš pētnieks, saglabājot protokolus, programmas un skriptus

Datu dokumentēšana

- **Metodoloģijas apraksti** (pētījuma dizaina apraksts, eksperimenta protokols, datu vākšanas procedūras)
- **Datu vākšanas instrumenti** (anketas, interviju ceļveži, kodu grāmatas, klasifikācijas shēmas)
- **Ētikas dokumenti** (daļbnieku informācijas lapas, informētās piekrišanas veidlapas)
- **Metadati** (ReadMe datne, kodu grāmata, saites uz pētījuma pārskatiem un publikācijām)

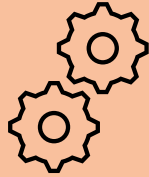
Pseudonimizācija vs. anonimizācija

	Pseudonimizācija	Anonimizācija
Definīcija	Personas datu apstrāde tā, lai tos vairs nevarētu tieši saistīt ar konkrētu personu bez papildu informācijas.	Personas datu apstrāde tā, lai datu subjektu vairs nebūtu iespējams identificēt nekādā veidā.
Atgriezeniskums	Atgriezeniska (var atjaunot saikni ar identitāti).	Neatgriezeniska (saikni ar identitāti nevar atjaunot).
Juridiskais statuss	Joprojām tiek uzskatīti par personas datiem.	Vairs netiek uzskatīti par personas datiem.
VDAR piemērošana	VDAR pilnībā attiecas.	VDAR vairs neattiecas.
Drošības līmenis	Zemāks nekā anonimizētiem datiem (nepieciešami papildu drošības pasākumi).	Augstāks nekā pseudonimizētiem datiem (identifikācija teorētiski nav iespējama).

Datu apstrāde un analīze

1010
1010

**Jēldatu
saglabāšana**



**Rīki, kas nodrošina
skriptu izveidi**
(R, Python...)



**Detalizēta
dokumentācija**
*(metodes,
parametri un rīki)*

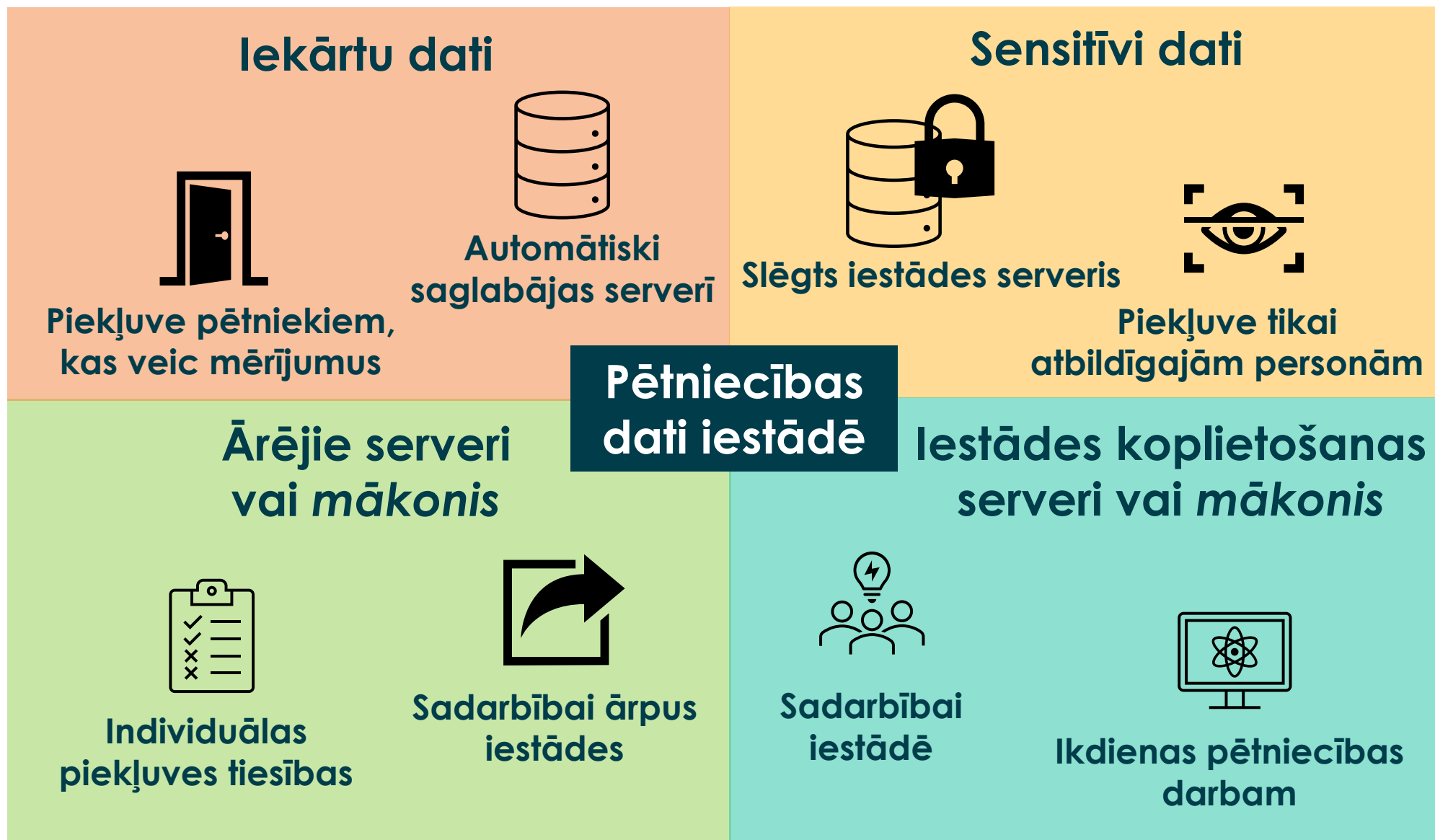


**Reproducējamības
pārbaude**
*(vai kāds var atkārtot
analīzi?)*



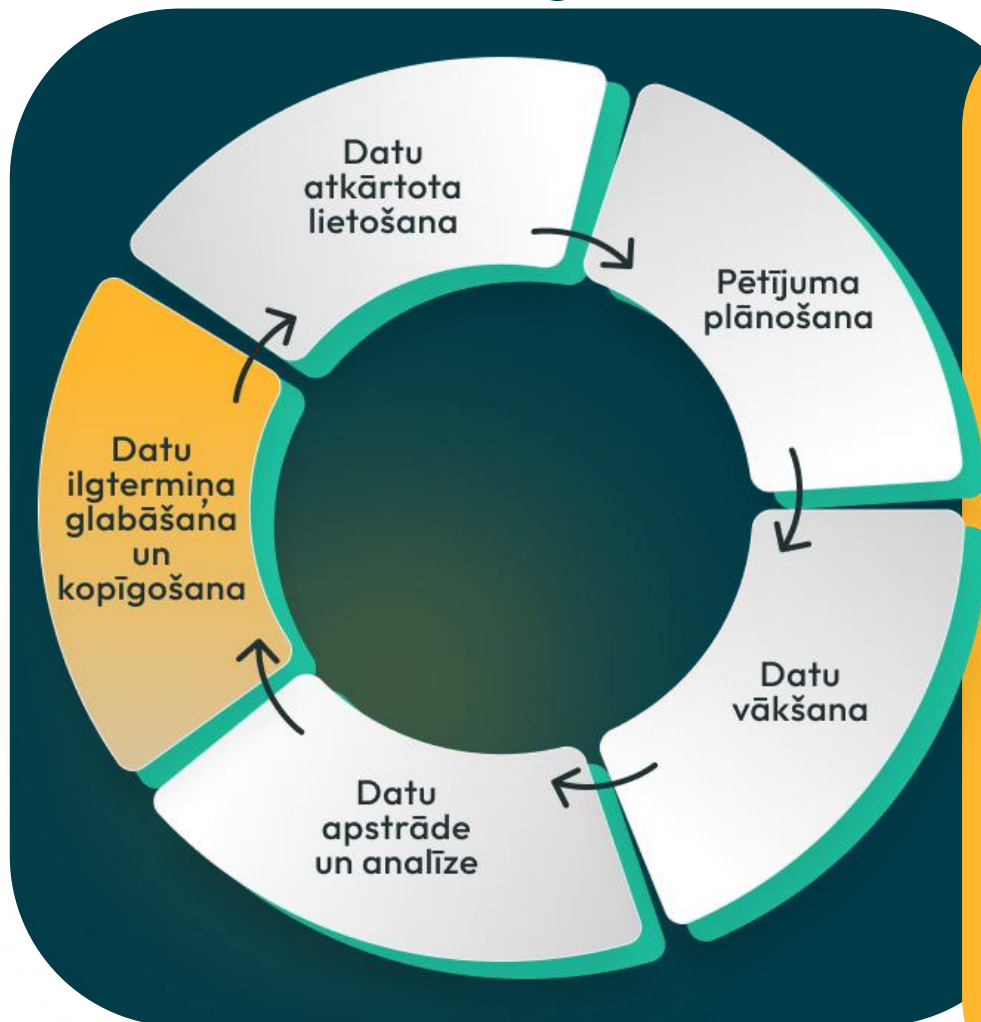
- Apstrāde un analīze ir pārbaudāma un reproducējama
- Rezultāti ir caurspīdīgi un uzticami

Iestādes infrastruktūra



Darbinieku datori izmantojami tikai datu apstrādei un analīzei. Tie nav paredzēti datu glabāšanai.

Ilgtermiņa glabāšana un kopīgošana



Publicējamo datu atlase – prasības, atkalizmantošanas potenciālu

Repozitorija izvēle – vispārīgais (*DataverseLV*), nozares, specifiskais, atbilstoši prasībām

Datu aprakstīšana – metadatu ieraksts repozitorijā, dokumentācija, lai nodrošinātu datu atrodamību un saprotamību

Licences un piekļuve – izmantošanas nosacījumi

Deponēšana (pašarhivēšana) – datu kopas, datņu augšupielāde repozitorijā

Atbalstāmie datņu formāti

Datnes veids	Vēlamie formāti	Pieņemamie formāti	Nevēlamie formāti
Teksta dokumenti	<ul style="list-style-type: none">PDF/A (.pdf)Markdown (.md)Unicode Text (.txt)	<ul style="list-style-type: none">RTF (.rtf)PDF (.pdf)OpenOffice XML (.docx)	<ul style="list-style-type: none">Microsoft Word (.doc)Google Docs (.gdoc)
Prezentācijas	<ul style="list-style-type: none">PDF/A (.pdf)	<ul style="list-style-type: none">OpenOffice PowerPoint XML (.pptx)OpenDocument Presentation (.odp/.sxi)	<ul style="list-style-type: none">Microsoft PowerPoint (.ppt)Google Slides
Datu tabulas	<ul style="list-style-type: none">CSV (.csv)Delimited text (.txt)TAB (.tab)	<ul style="list-style-type: none">XML Workbook (.xlsx)SPSS (.sav)R (.RData, .rds)	<ul style="list-style-type: none">Microsoft Excel (.xls)SPSS Portable (.por)Stata (.dta)Matlab (.mat)
Audio	<ul style="list-style-type: none">FLAC (.flac)MXF (.mxf)WAVE (.wav)	<ul style="list-style-type: none">AIFF (.aif)MPEG-4 Audio (.mp4)MP3 (.mp3)	<ul style="list-style-type: none">AAC (.aac, .m4a)WMA (.wma)
Attēli	<ul style="list-style-type: none">TIFF (.tif, .tiff)PNG (.png)SVG (.svg)	<ul style="list-style-type: none">JPEG (.jpg, .jpeg)GIF (.gif)BMP (.bmp)	<ul style="list-style-type: none">Adobe Photoshop (.psd)Apple Picture (.pct)

Tabulāro datu kvalitāte

1. Katram failam ir **atbilstošs** un **saprotams** nosaukums
2. Katrā failā ir tikai **viena** tabula
3. Katrai **kollonai** tikai viens mainīgais
4. Katrā **rindā** tikai viens paraugs
5. Katrā **šūnā** tikai viena vērtība un tā visur ir konsekventa (~~Mērijums, mērijums, MĒRĪJUMS...~~)
6. Šūnas nav formatētas (nav krāsojumu, fonta maiņa...)
7. Pirmā rindā ir kolonnu nosaukumi, kuri nesatur īpašas rakstzīmēs un nesākas ar cipars
8. **Tukšās vērtības** apzīmētas ar NA, NULL vai cits

country	year	cases	population
Afghanistan	1999	1645	19987071
Afghanistan	2000	1666	20095360
Brazil	1999	37737	172006362
Brazil	2000	80488	174004898
China	1999	210258	1272015272
China	2000	210066	1280048583

Mainīgie

country	year	cases	population
Afghanistan	1999	1645	19987071
Afghanistan	2000	1666	20095360
Brazil	1999	37737	172006362
Brazil	2000	80488	174004898
China	1999	210258	1272015272
China	2000	210066	1280048583

Novērojumi

country	year	cases	population
Afghanistan	99	75	987071
Afghanistan	00	66	095360
Brazil	99	737	006362
Brazil	00	488	004898
China	99	258	015272
China	00	066	048583

Vērtības

Kodu grāmata

Kolonnas nosaukums	Apraksts	Datu tips	Mērvienība	Piemērs	Piezīmes
ID	Ieraksta identifikators	Integer	–	1	Unikāls katram ierakstam
Datums	Novērojuma datums	Date	YYYY-MM-DD	2025-10-01	ISO 8601 formāts
Temp_C	Temperatūra gaisā	Float	°C	18.5	Mērīts ar sensoru X
Mitrums_proc	Relatīvais mitrums	Integer	%	65	Noapaļots līdz vesalam
Stacija	Mērījumu stacijas nosaukums	String	–	Jelgava	Bez diakritiskām zīmēm
Piezīmes	Papildu piezīmes par apstākļiem	String	–	Lietus	Brīvā teksta lauks

Repozitorija izvēle



Datu aprakstīšana un metadati

- Metadati – dati par datiem:
 - standartizēti un strukturēti
 - cilvēkiem un mašīnām saprotamā valodā
- Nozares metadatu standarti
- Kur norāda metadatus:
 - metadatu ieraksti (lauki) repozitorijā
 - datu/pētījuma dokumentācijā (piemēram, *README* datnē, kas pievienota datiem)
- Dokumentācija īpaši nepieciešama datu atkārtotas izmantošanas gadījumā

Dokumentācija

Metadati

- Nosaukums
- Autori
- Apraksts (tostarp citi pētījuma rezultāti ar pastāvīgo identifikatoru)
- Atslēgvārdi
- Finansējums
- Kontaktinformācija

Piekļuve un licence

Piekļuve:

- atvērtā, slēgtā, ierobežotā
- embargo (pieejami pēc noteikta laika)

Licence

- ko citi drīkst darīt ar datiem
- pēc iespējas atvērtāka izvēle – vislabāk **CC BY**

VISVAIRĀK
ATVĒRTĀS

CREATIVE COMMONS LICENCES



CCO PUBLIC DOMAIN DEDICATION

Norāda uz pilnīgu atteikšanos no autortiesībām. Datu kopa uzskatāma par sabiedrības īpašumu un brīvi izmantojama bez ierobežojumiem



CC BY ATTRIBUTION

Jānorāda atsauce uz autoru. Tad atļauta brīva datu kopas izmantošana, tajā skaitā arī komerciāliem nolūkiem



CC BY-SA ATTRIBUTION-SHAREALIKE

Jānorāda atsauce uz autoru un visi jaunie darbi tiek licencēti ar tādiem pašiem noteikumiem. Atļauta jebkāda datu kopas izmantošana, arī komerciāliem nolūkiem



CC BY-ND ATTRIBUTION-NODERIVS

Jānorāda atsauce uz autoru un atļauts izmantot datu kopu tikai neizmainītā veidā, arī komerciāliem nolūkiem



CC BY-NC ATTRIBUTION-NONCOMMERCIAL

Jānorāda atsauce uz autoru un atļauta izmantošana tikai nekomerciāliem nolūkiem



CC BY-NC-SA ATTRIBUTION-NONCOMMERCIAL-SHAREALIKE

Jānorāda atsauce uz autoru, atļauta izmantošana tikai nekomerciāliem nolūkiem, un visi jaunie darbi tiek licencēti ar tādiem pašiem noteikumiem

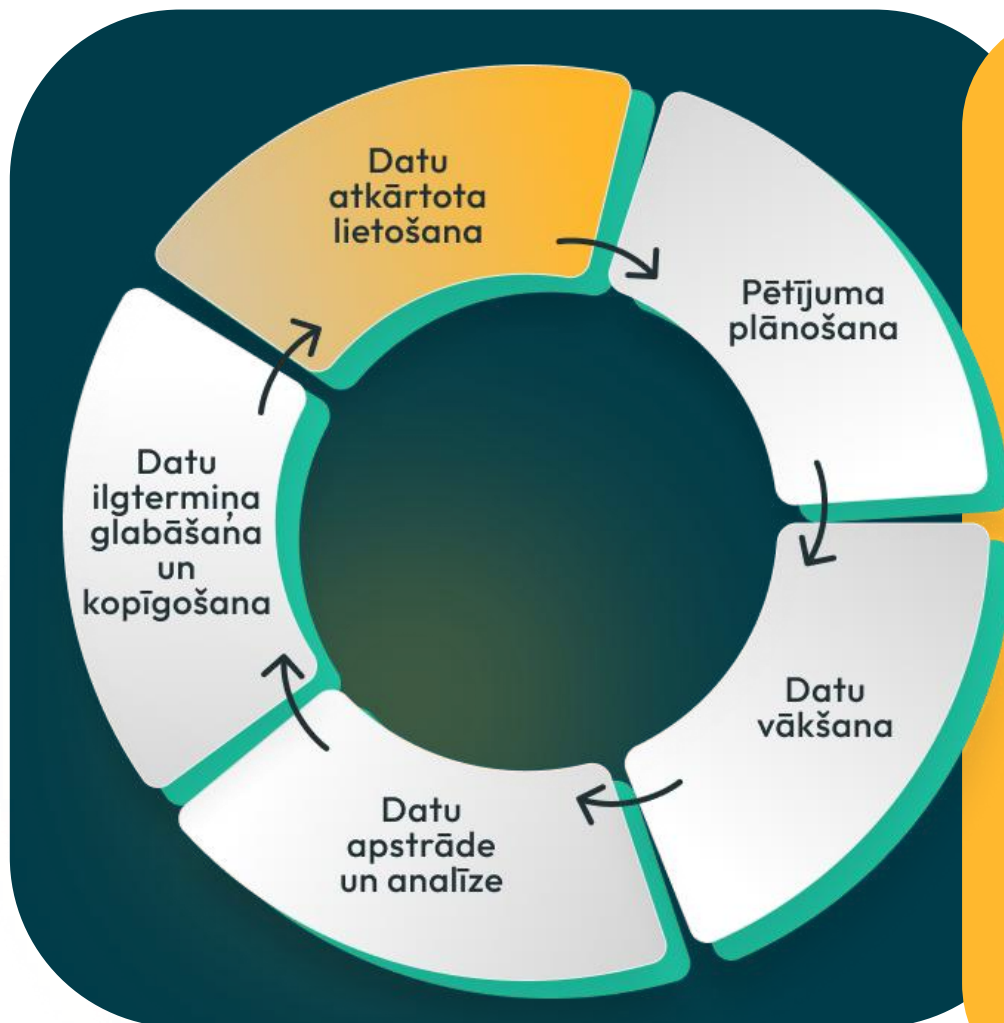


CC BY-NC-ND ATTRIBUTION-NONCOMMERCIAL-NODERIVS

Jānorāda atsauce uz autoru, atļauta darba izmantošana tikai nekomerciāliem nolūkiem un neizmainītā veidā

VISMAZĀK
ATVĒRTĀS

Datu atkārtota izmantošana



Datu atrašana – gan repozitorijos, gan speciālos meklētājos, kuros atrodami repozitoriju dati

Kad atrasti – metadatu, README un dokumentācijas izpēte, lai saprastu vai tie ir atbilstoši

Izmantošanas nosacījumi - piekļuves iespējas un izmantošanas nosacījumi, t.sk. licences

Atsauces – izmantojot datus, jāpievieno norādes (paraugs nereti pievienots repozitorija metadatu ierakstā), iekļauj arī DOI

Kāpēc padarīt datus pieejamus repozitorijā?



Atvērtās zinātnes principu ieviešana praksē



Atbilstība **finansētāju** prasībām



Atbilstība **zinātnisko žurnālu** prasībām un rekomendācijām



Lielāka **citējamība** un ietekme



Jaunas **sadarbības** iespējas

Paldies par uzmanību!

Materiāls izstrādāts projekta “Atbalsts atvērtās zinātnes ieviešanai praksē, kā arī izveidoti risinājumi zinātnes datu koplietošanai un dalībai ES atvērtajā zinātnes mākonī” ietvaros (ANM projekta Nr. 2.1.3.1.i) ar Eiropas Savienības Atvērto zinātnes fonda un Latvijas valsts finansiālo atbalstu



Finansē
Eiropas Savienība
NextGenerationEU



Nacionālais
attīstības plāns



Latvijas
Biozinātņu un tehnoloģiju
universitāte



RĪGAS TEHNISKĀ
UNIVERSITĀTE



LATVIJAS
UNIVERSITĀTE

RSU



VPC