

Ieskats datu pārvaldībā

Gita Rozenberga

Nacionālais atvērtās piekļuves dienests

Latvijas Universitātes bibliotēka

Prezentācijā...

Kas ir pētniecības datu pārvaldības ?

Pētniecības datu pārvaldības dzīves cikls

Pētniecības dati

Pētniecības dati ir fiksēta dažāda veida informācija: fakti, novērojumi, attēli, video/audio ieraksti, datorprogrammu rezultāti, mērījumi un cita veida pieredze, uz kuras balstās argumenti, teorijas, testi, hipotēzes un citi pētniecības rezultāti.

Tie var būt jēldati, dati, kuriem veikta pirmapstrāde jeb “tīrīšana”, un pilnībā apstrādāti dati, kas saglabāti jebkurā formātā un informācijas nesējā.

Veids, kā dati tiek organizēti un arhivēti, ir ļoti svarīgs, lai pētījumus varētu atkārtot un pārbaudīt iegūtos rezultātus.

Pētniecības datu pārvaldība (1)

Pētniecības datu pārvaldība ietver procesus un darbības, kas vajadzīgas, lai pārvaldītu datus visa pētniecības dzīves cikla laikā, gan pašreizējiem, gan turpmākiem pētījumiem un interesentu vajadzībām.

Svarīga pētījuma dzīves cikla komponente ir pētniecības datu pārvaldības plāns.



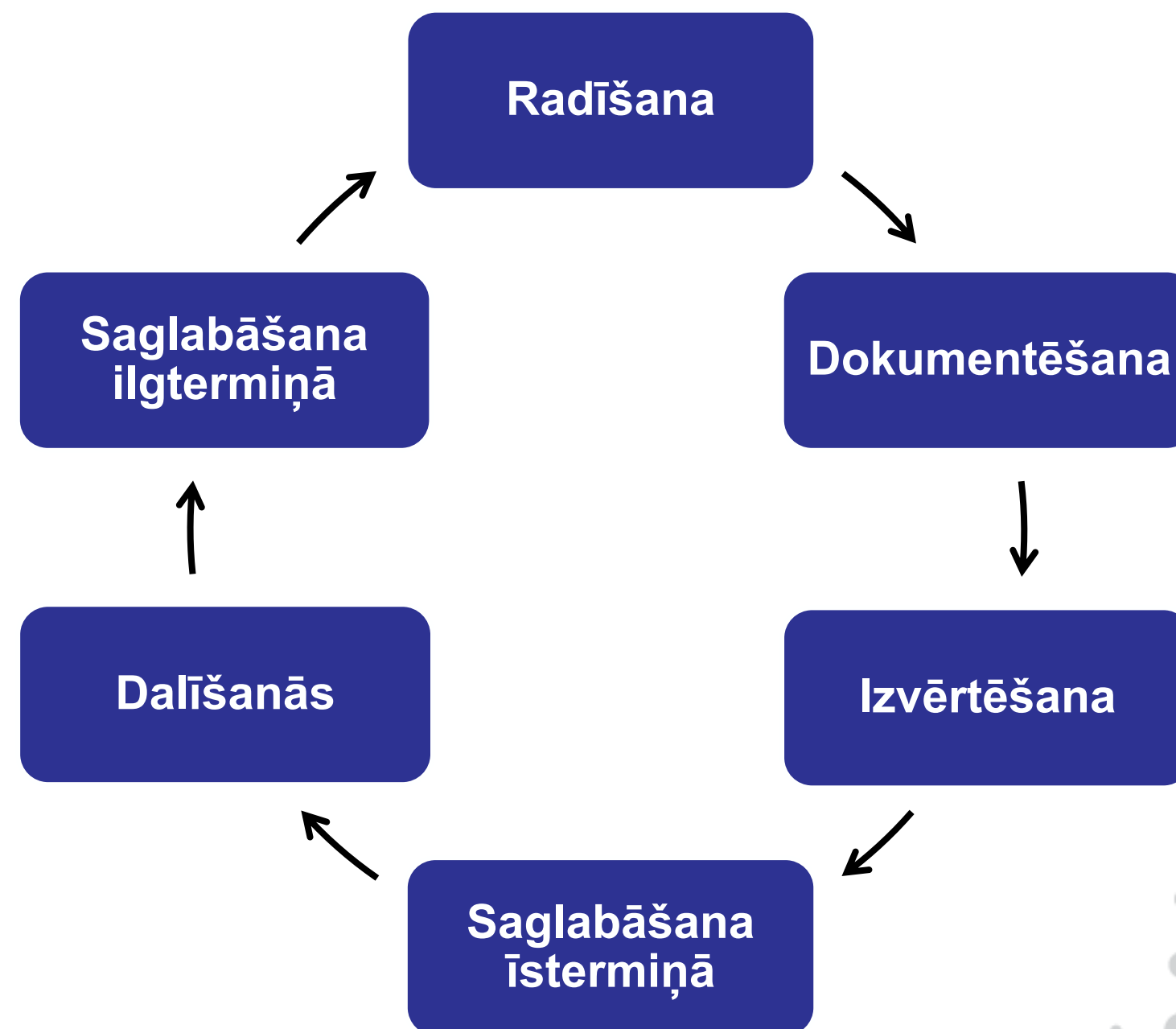
Pētniecības datu pārvaldība (2)

Koncepts «pētniecības datu pārvaldība» nozīmē:

- Skaidrojums, ko iesākt ar datiem, kā organizēt un saglabāt pētniecības datus
- Uzmanības pievēršana tehniskiem, organizatoriskiem, strukturāliem, likumdošanas un ilgtspējības aspektiem
- Skaidra struktūra kā dati tiks pārvaldīti
- Papildus darba uzdevums (datu pārvaldības plāna izveide) pētījuma agrā sākuma stadijā, kas ilgtermiņā atmaksājas

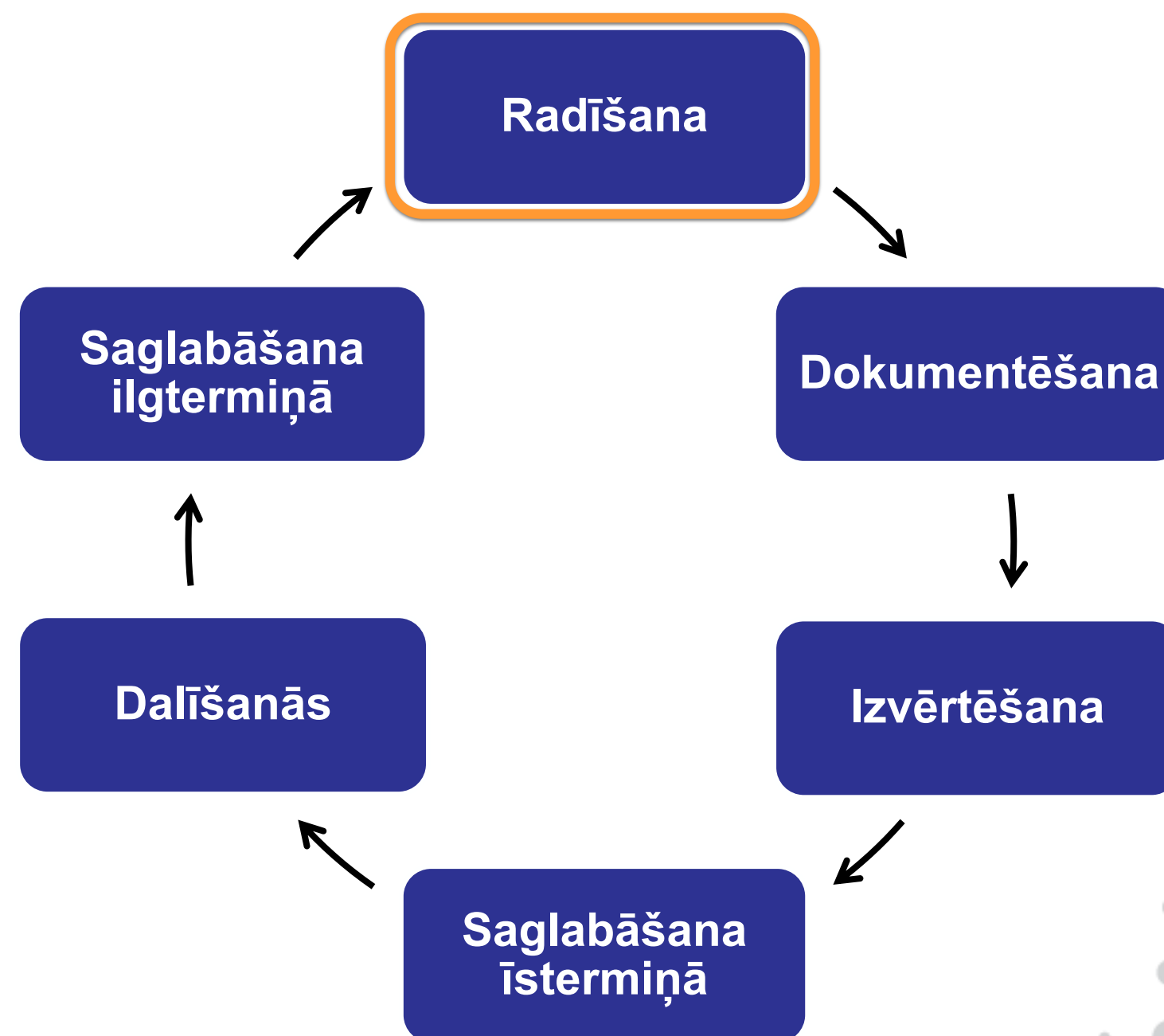
Pētniecības datu pārvaldības dzīves cikls (1)

Tas ir aktīvs process



Pētniecības datu pārvaldības dzīves cikls (2)

- Respondenta informētās piekrišanas veidlapas, licences un vienošanās līgumi, kas dod atļauju dalīties ar datiem.
- Datņu formāti, kas nodrošina pēc iespējas lielāku atvērtību
- Failu nosaukšanas kārtība un pētījuma projekta grupas informētība par to
- Metadatu piefiksēšana un datu kopu dokumentēšana (laicīgi)



Dalības piekrišanas veidlapas

Respondenta **informētās piekrišanas veidlapās** jāietver:

- Pētījuma mērķis un līdzdalības raksturs
- Skaidrojums, kā tiks saglabāta konfidencialitāte
- Dažādu dalības, publicēšanas un datu kopīgošanas piekrišanas nosacījumu iespējas
- Informācija par datu ilgtermiņa saglabāšanu
- Datu izmantošana turpmākai izpētei, mācībām u.c.
- Piemērs: https://www.ed.ac.uk/files/imports/fileManager/Interview_Consent_Form.pdf

Datu centri var atteikties pieņemt datus, ja nav izmantotas un savāktas informētās piekrišanas veidlapas

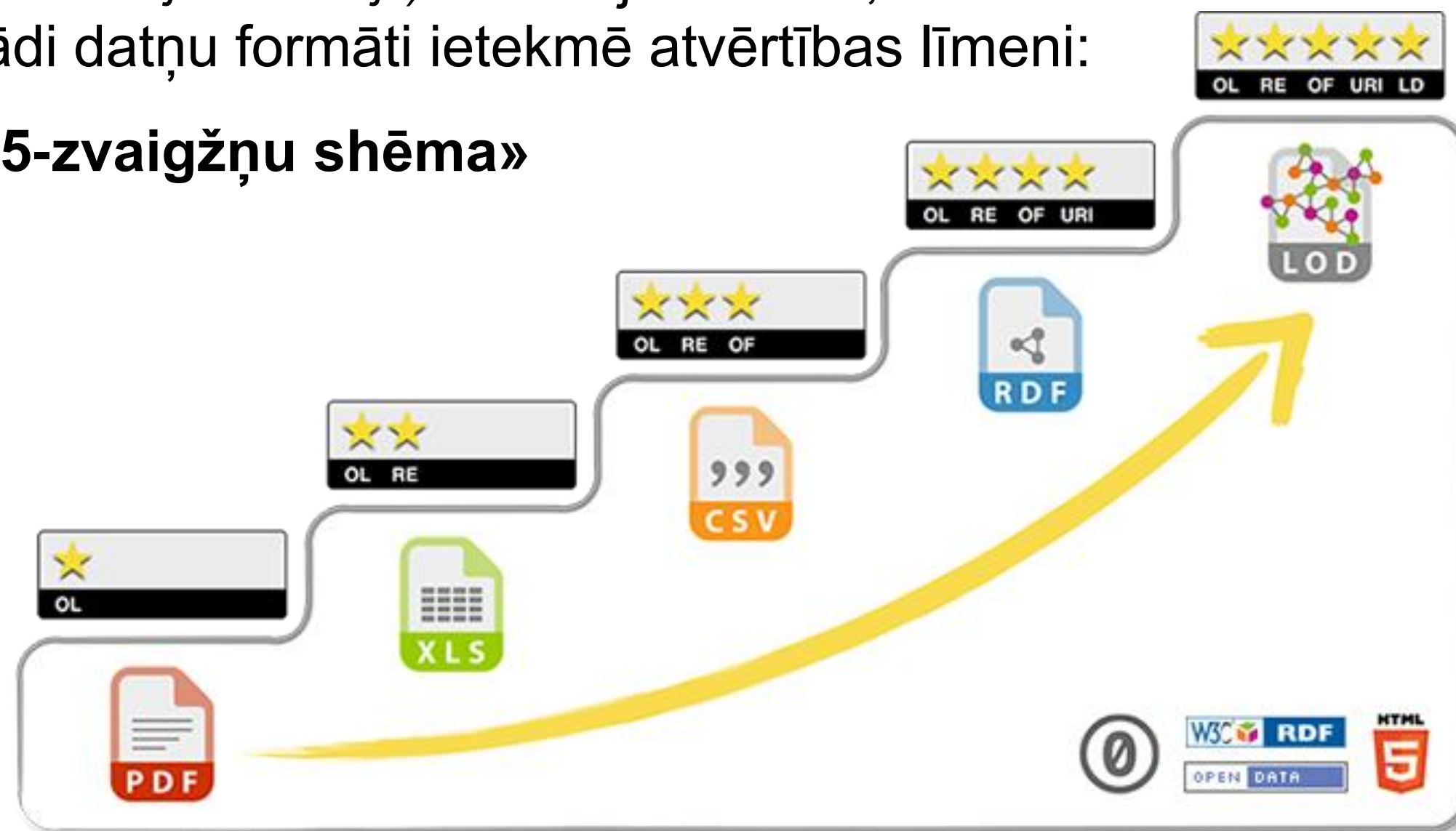
Datņu formāti (1)

- Dažādi **datņu formāti** ir **noderīgi dažādām vajadzībām**
- Datu saglabāšanai noderīgāki ir atvērtie un bezzudumu formāti
- Lai gan patentēto un saspiesto datņu formāti ir plaši izmantoti, tie nav labi noderīgi ilgtermiņa glabāšanai
- Datnes vajag **konvertēt uz nepieciešamajiem formātiem** atbilstoši pētījuma procesa vajadzībām (piem., analīzei var būt nepieciešami specifiski formāti, savukārt, saglabāšanai citi)

Datņu formāti (2)

Tims Bērnss-Lī (*vispasaules tīmekļa izgudrotājs un vispasaules tīmekļa konsorcijs pašreizējais vadītājs*) izveidojis shēmu, kas labi ilustrē, kā dažādi datņu formāti ietekmē atvērtības līmeni:

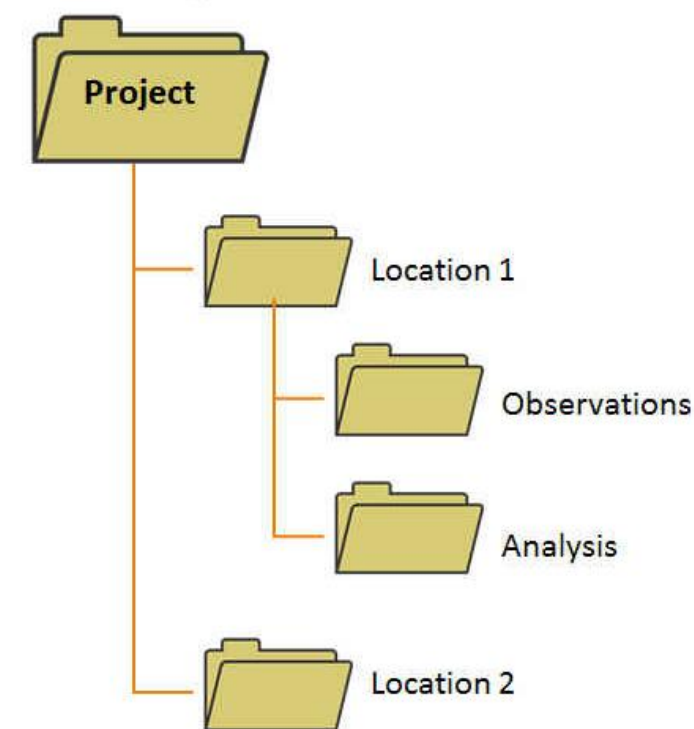
«Atvērto datu 5-zvaigžņu shēma»



Datņu organizēšana

Pamatnoteikumi:

- **Panākt līdzsvaru datņu nosaukumu elementos** (Pārāk maz elementu rada neskaidrību. Pārāk daudz – apgrūtina meklēšanu un izpratni);
- Izmantot **jēgpilnus** saīsinājumus;
- Vienoties par **versiju apzīmējumiem**;
- Vēlams neizmantot simbolus, kas nav burti vai cipari;
- Neizmantot «tukšumu» (tā vietā labāk lietot apakšsvītru);
- Datnes tiks sakārtotas, sākot ar pirmajiem nosaukuma elementiem, tāpēc jāsāk ar vispārīgākiem un jānoslēdz ar specifiskākiem elementiem. (Datuma elementam vienmēr jāsākas ar gadu, tad mēnesi un datumu, lai datnes tiktu sakārtotas hronoloģiski);
- **Pierakstīt savus lēmumus par datņu nosaukšanas kārtību**;
- **Informēt kolēģus par kopīgajā projektā izmantoto datņu nosaukšanas kārtību** un rūpēties, lai līdzautori zinātu, kā to izmantot.

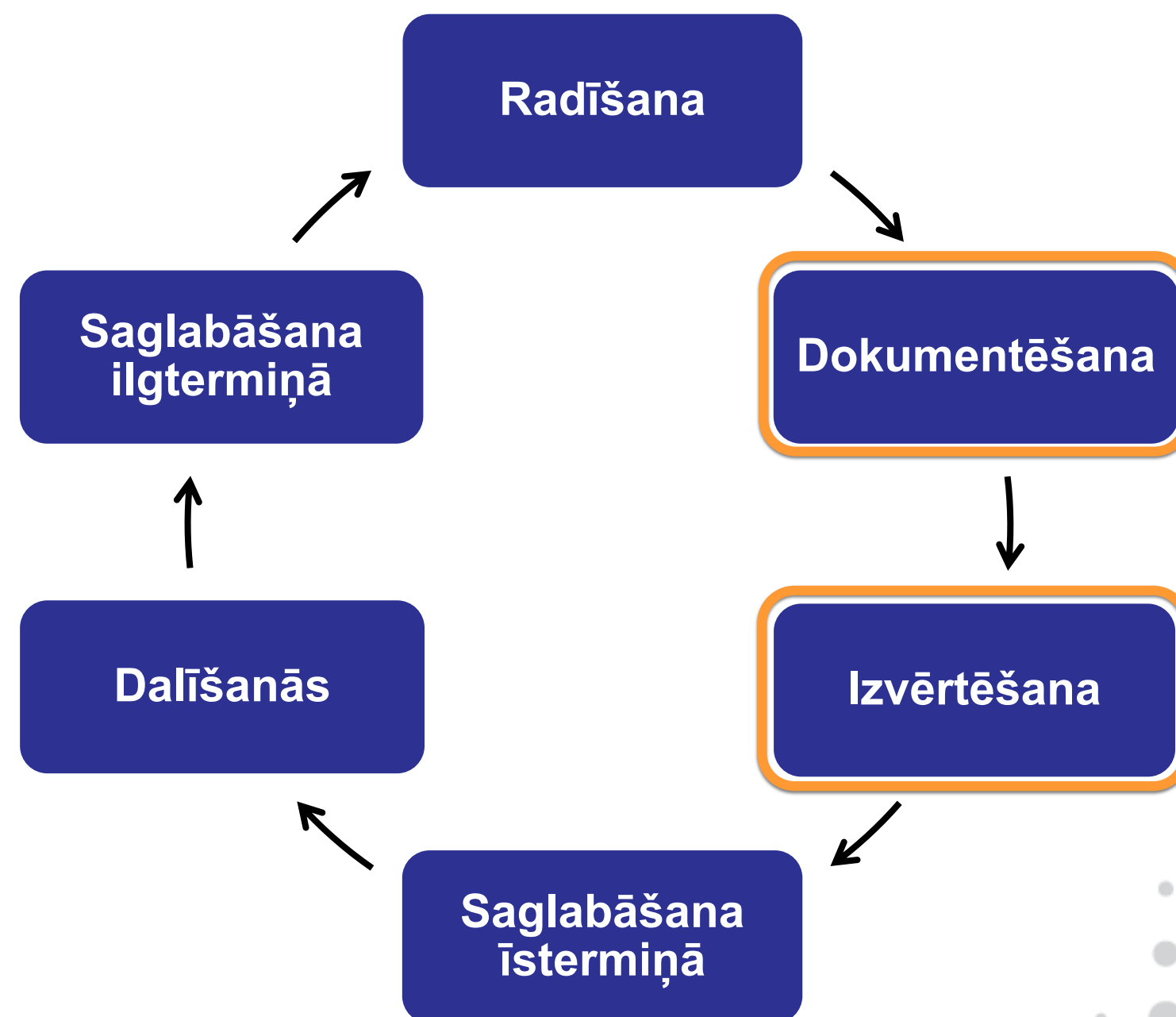


AVOTS: <https://libguides.princeton.edu/c.php?g=102546&p=930626>

Pētniecības datu pārvaldības dzīves cikls (3)

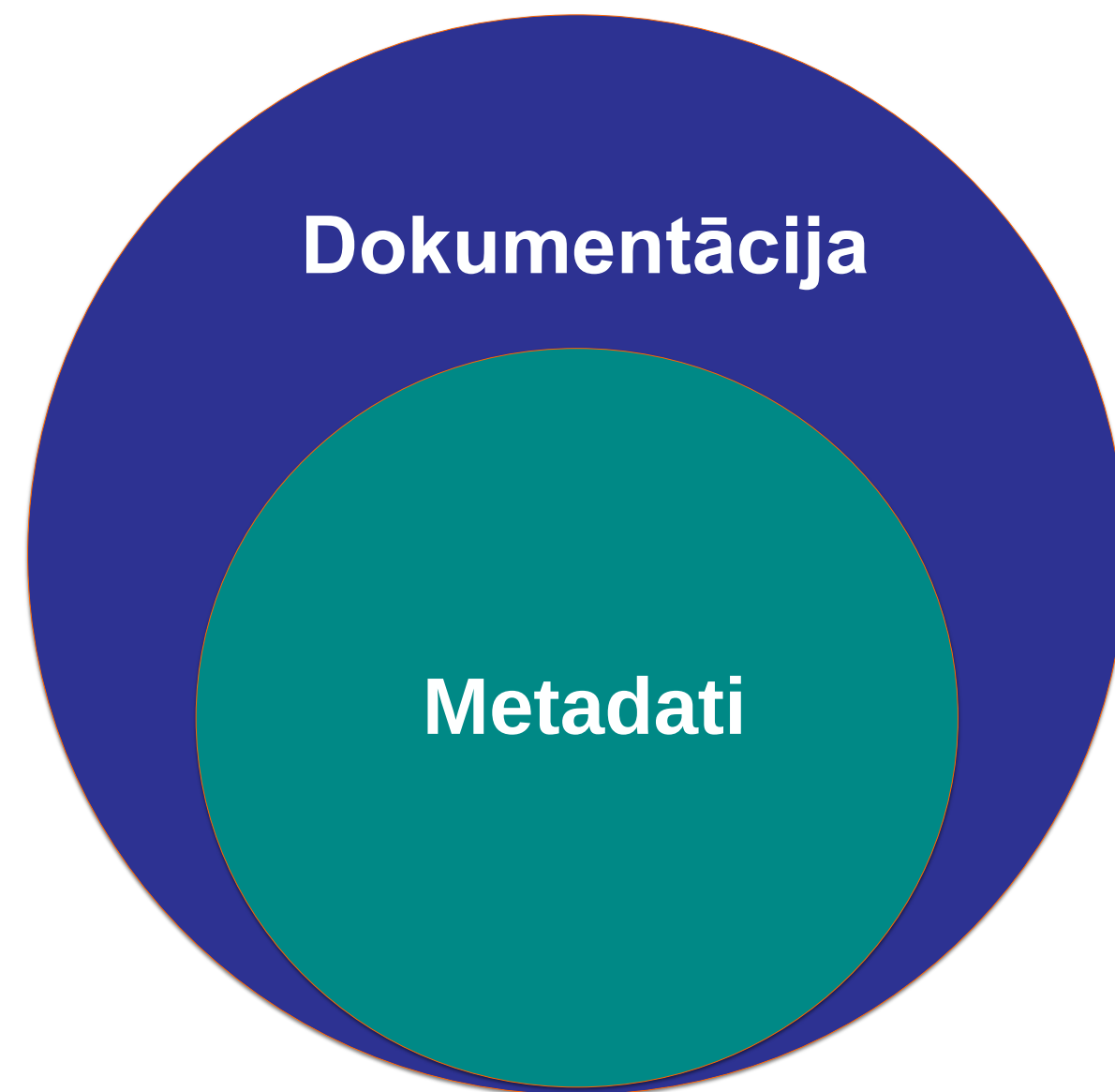
Pārdomāt, kas nepieciešams, lai spētu novērtēt, saprast un atkārtoti izmantot datus:

- Kāpēc dati tika radīti?
- Kādi procesi un kā tiem tika veikti?
- Vai tika izveidots kods, lai tos analizētu? (arī jā saglabā un/vai jākopīgo)
- Sniegt plašāku kontekstu uzticamības veicināšanai



Metadati (1)

- Metadatu pazīmes:
 - Standartizēti
 - Strukturēti
 - Cilvēkiem un mašīnām saprotamā valodā (*Machine and human readable*)
- Sniedz konteksta informāciju, kas palīdz interpretēt
- Metadati palīdz atsaukties jeb citēt datus
- Dokumentācijas informācija nepieciešama atkārtotas izmantošanas gadījumā



Metadati (2)

Datu aprakstīšanas pieejas:

Pamata: nosaukums, datējums, ģeogrāfiskās atrašanās vietas, priekšmeti.

Savienojumi: pētnieki, līdzstrādnieki, saistītās publikācijas, tīmekļa vietnes, projekti un datu kopas.

Pieeja un tiesības: autortiesību licences, piekļuves un lietošanas ierobežojumi, embargo periodi.

Tehniskais: datņu formāts un izmērs, programmatūra, programmēšanas valoda.

Saglabāšana: uzglabāšanas vieta un formāts, saglabāšanas periodi.

- Dati par datiem
- Sastāv no atribūtu kopas
- Mašīnlasāmi, padarot datus atrodamus
- Palīdz novērst neatbilstošu izmantošanu
- Efektīvi ir izmantot savas nozares metadatu standartus

Metadati (3)

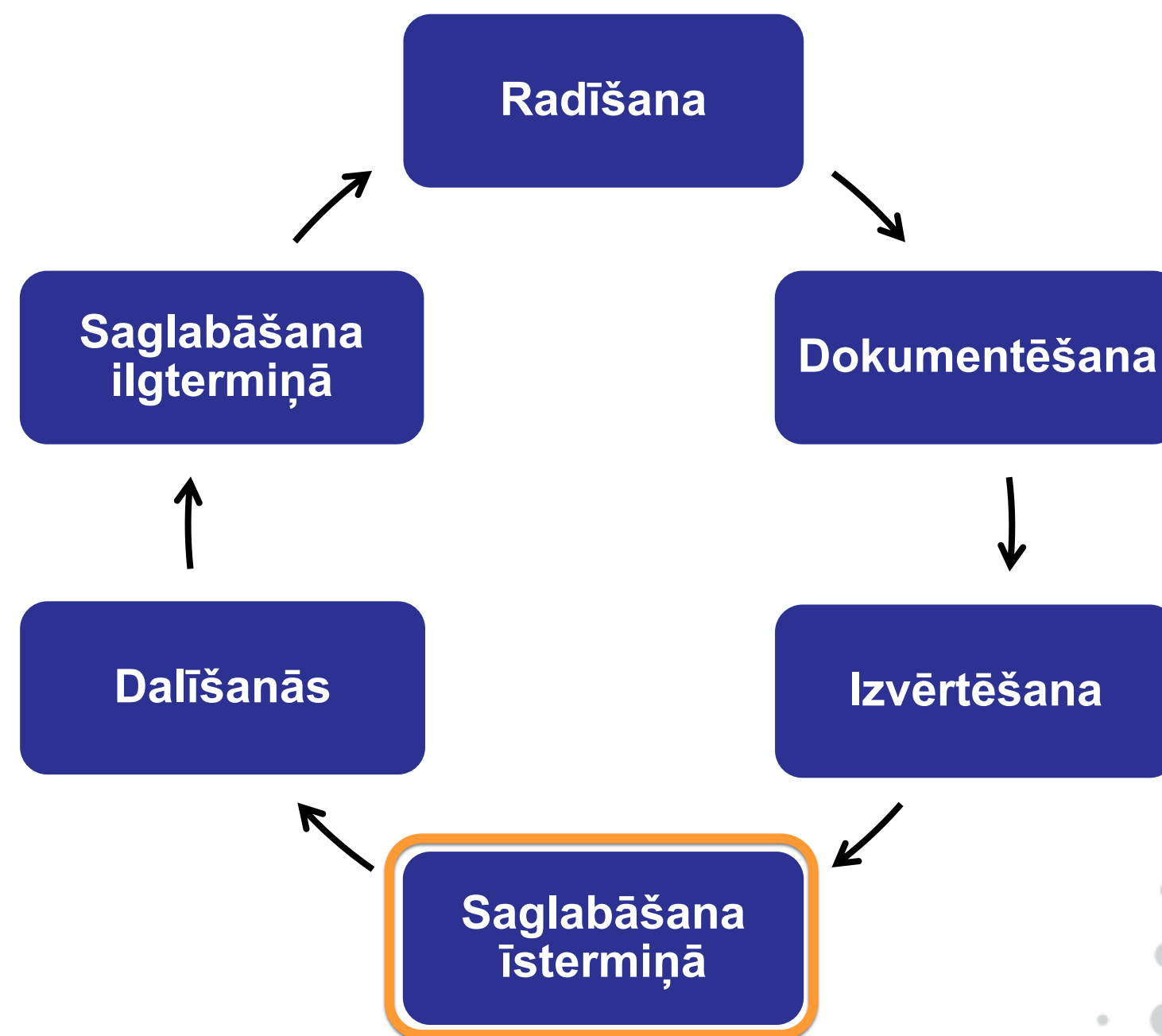
Metadati un atvērtība

- Pieejami metadati par pētniecības datiem ir jau solis tuvāk atvērtai zinātnei
- Atbilstoši standartiem veidoti metadati ir jau nākamais solis tuvāk atvērtai zinātnei
- Atbilstoši konkrētas zinātņu nozares metadatu standartiem veidoti metadati ir jau nākamais solis vēl tuvāk atvērtai zinātnei

Pētniecības datu pārvaldības dzīves cikls (4)

Pārdomāt, kur uzglabās datnes īstermiņā? Vai izvēlētais risinājums ir pietiekoši drošs un kvalitatīvs?

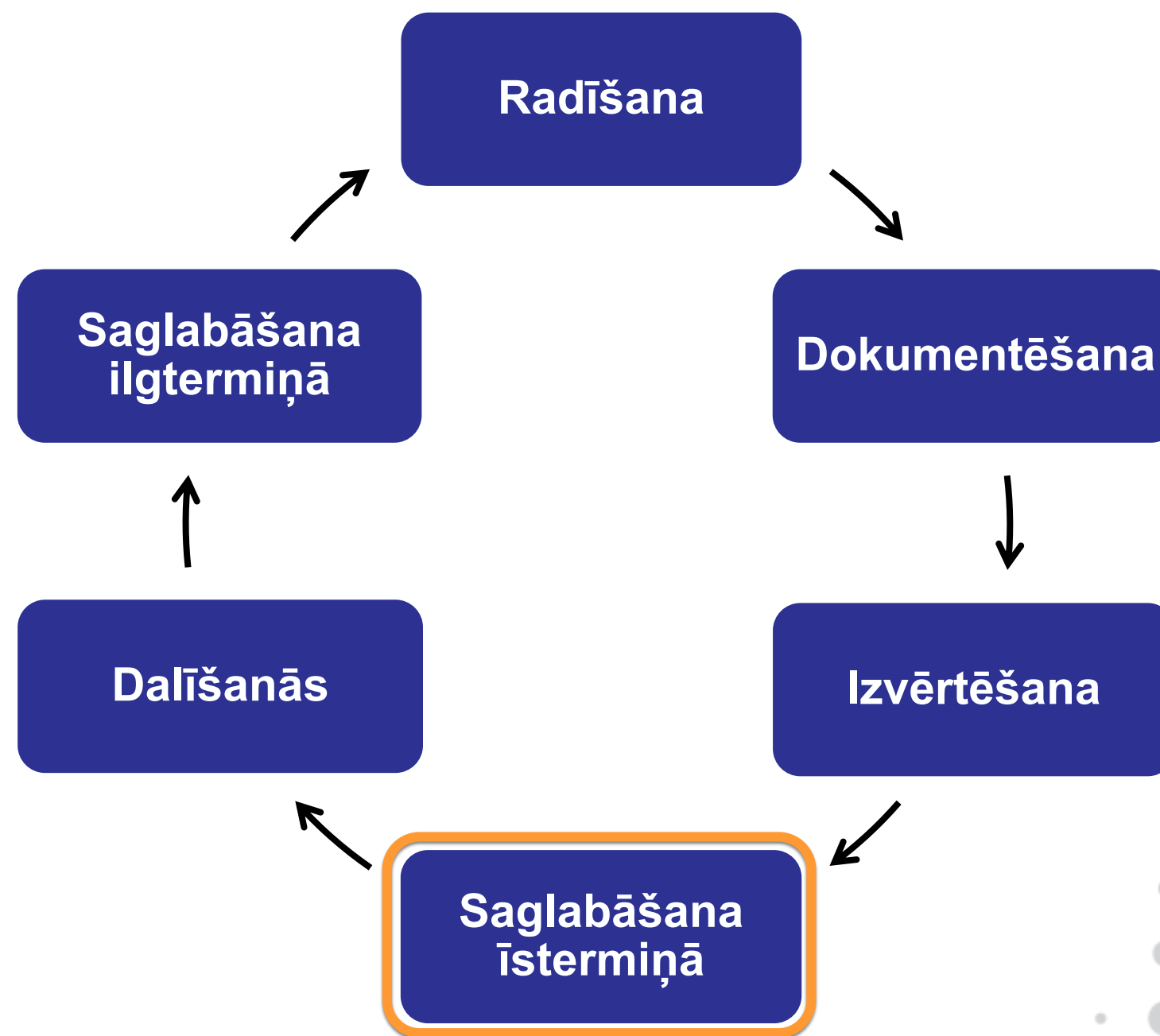
- Personīgā ierīce (piem., dators, USB zibatmiņa) ir ļoti riskanta vieta, kur glabāt
- Pārstāvētās institūcijas vai departamenta ārējie cietie diski un serveri
- Trešo pušu pakalpojumi, piem., Dropbox, Google Drive, OneDrive un citi mākoņpakalpojumi (nav droši, jo ir ar komerciālu ievirzi)



Pētniecības datu pārvaldības dzīves cikls (5)

Lēmumu izstrādājot jāuzdod jautājumi:

- Cik sensitīvi dati ir?
- Cik apjomīga un jaudīga krātuve nepieciešama?
- Kam un kad būs nepieciešama pieeja datiem?



Tiek ieteikts
izmantot *ownCloud*
pakalpojumus

Datu rezerves kopiju veidošana un saglabāšana ilgtermiņā

ATŠKIRĪGI PROCESI

Datu rezerves kopiju veidošana

- Regulāra datu pēdējās versijas kopēšana gadījumam, ja dati tiek iznīcināti vai pazūd
- Šīs datu kopijas tiek glabātas īsu termiņu
- Bieži pēc neregulāra grafika

Saglabāšana ilgtermiņā

- Sargāšanai no katastrofu riskiem
- Parasti ilgtermiņā saglabā datu kopu gala versijas
- Bieži tiek veikta projekta beigās vai galveno posmu pabeigšanas laikā

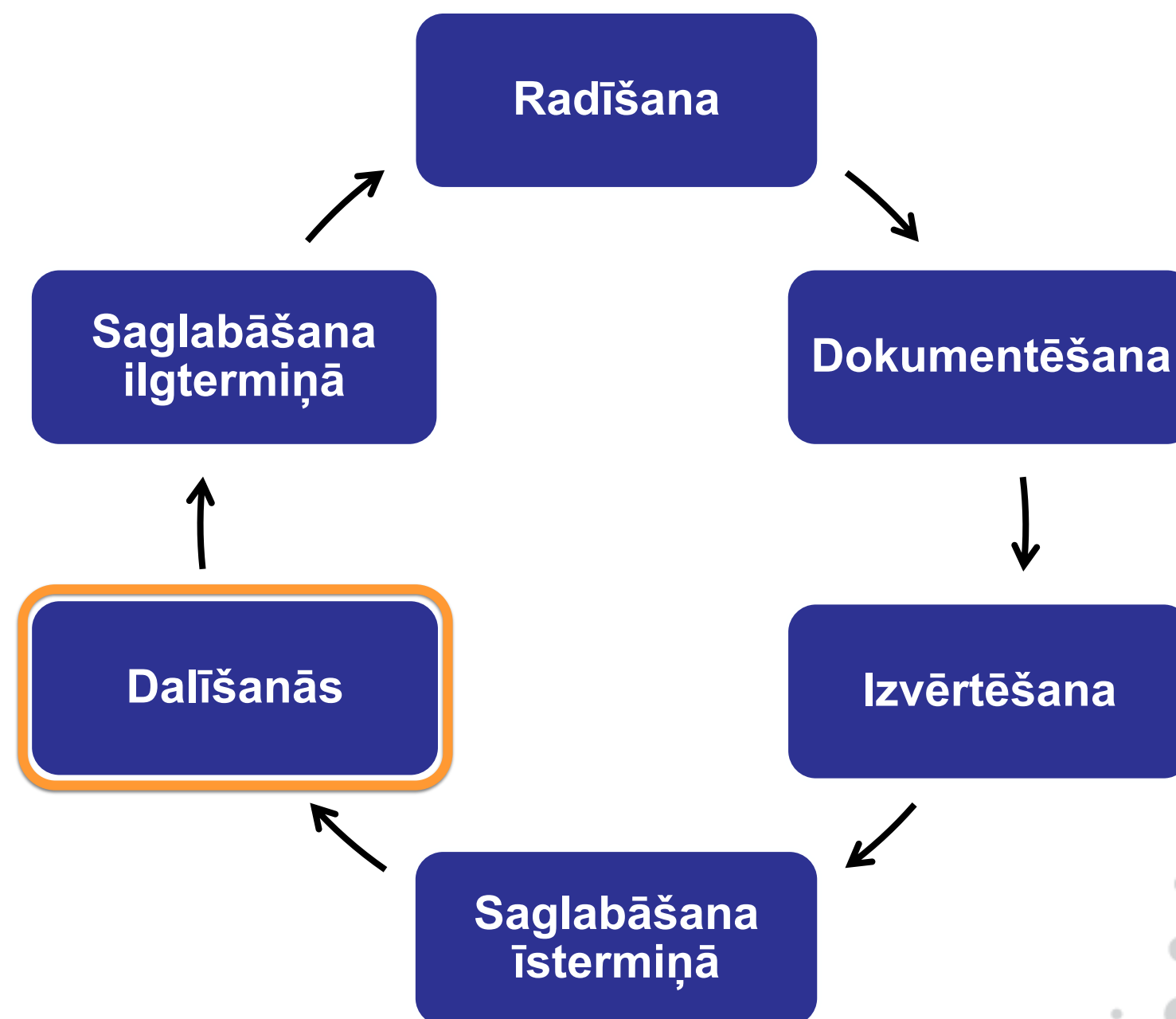


AVOTS:<https://www.digitaltrends.com/computing/keeping-data-safe-eternity/>

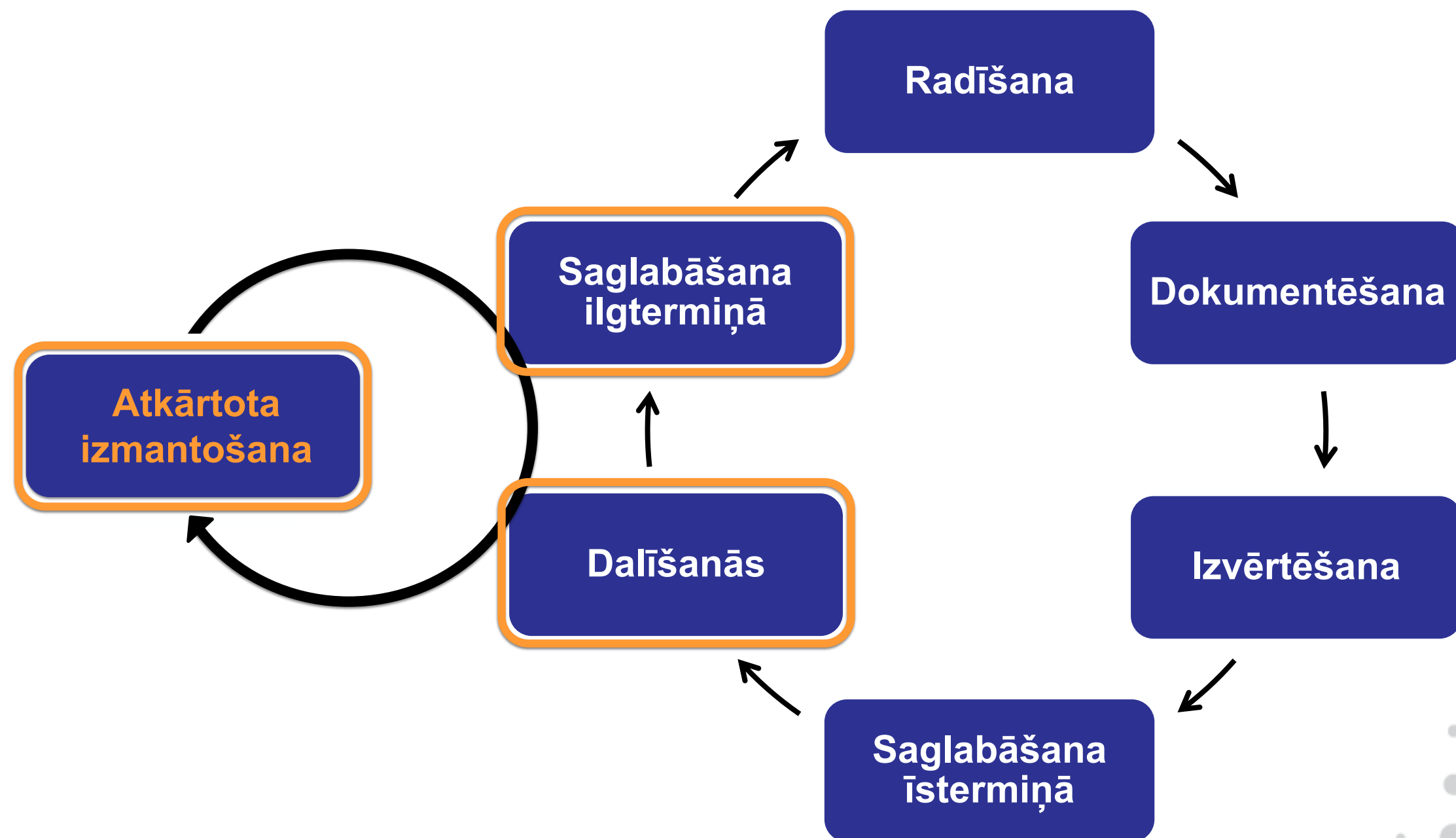
Pētniecības datu pārvaldības dzīves cikls (6)

Iedziļināties jautājumos un risinājumos, kā un uz kādiem nosacījumiem tiks nodrošināta pieeja datiem citiem interesentiem ārpus projekta?

- Izmantojot licences, pievienot informāciju, ko drīkstēs darīt ar datiem.
- Jo atvērtāku licences veidu pievienos, jo vieglāka un plašāka būs datu izmantošana, tostarp ilgtermiņā.



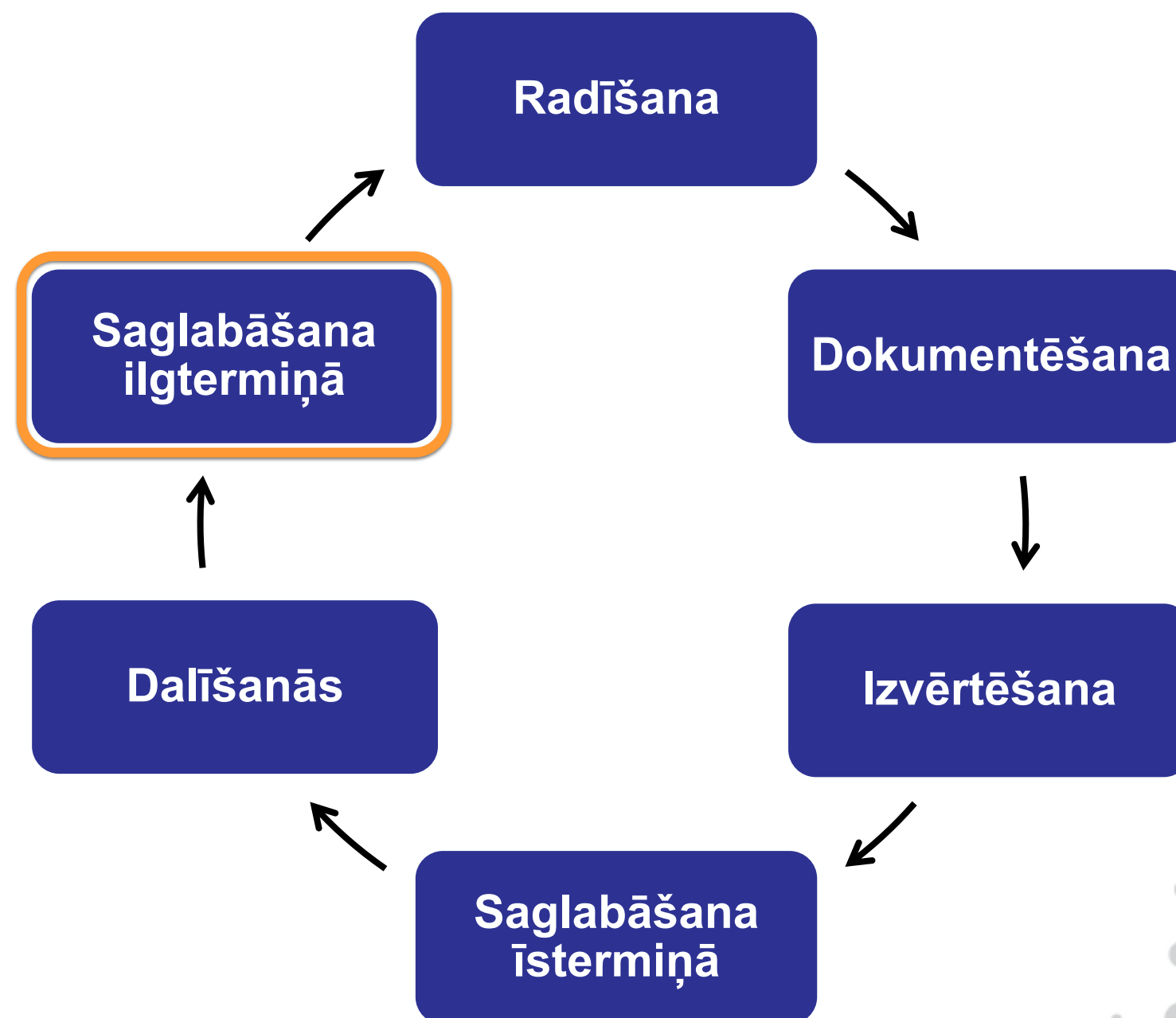
Pētniecības datu pārvaldības dzīves cikls (7)



Pētniecības datu pārvaldības dzīves cikls (8)

Datu deponēšana repozitorijā
vai digitālā arhīvā

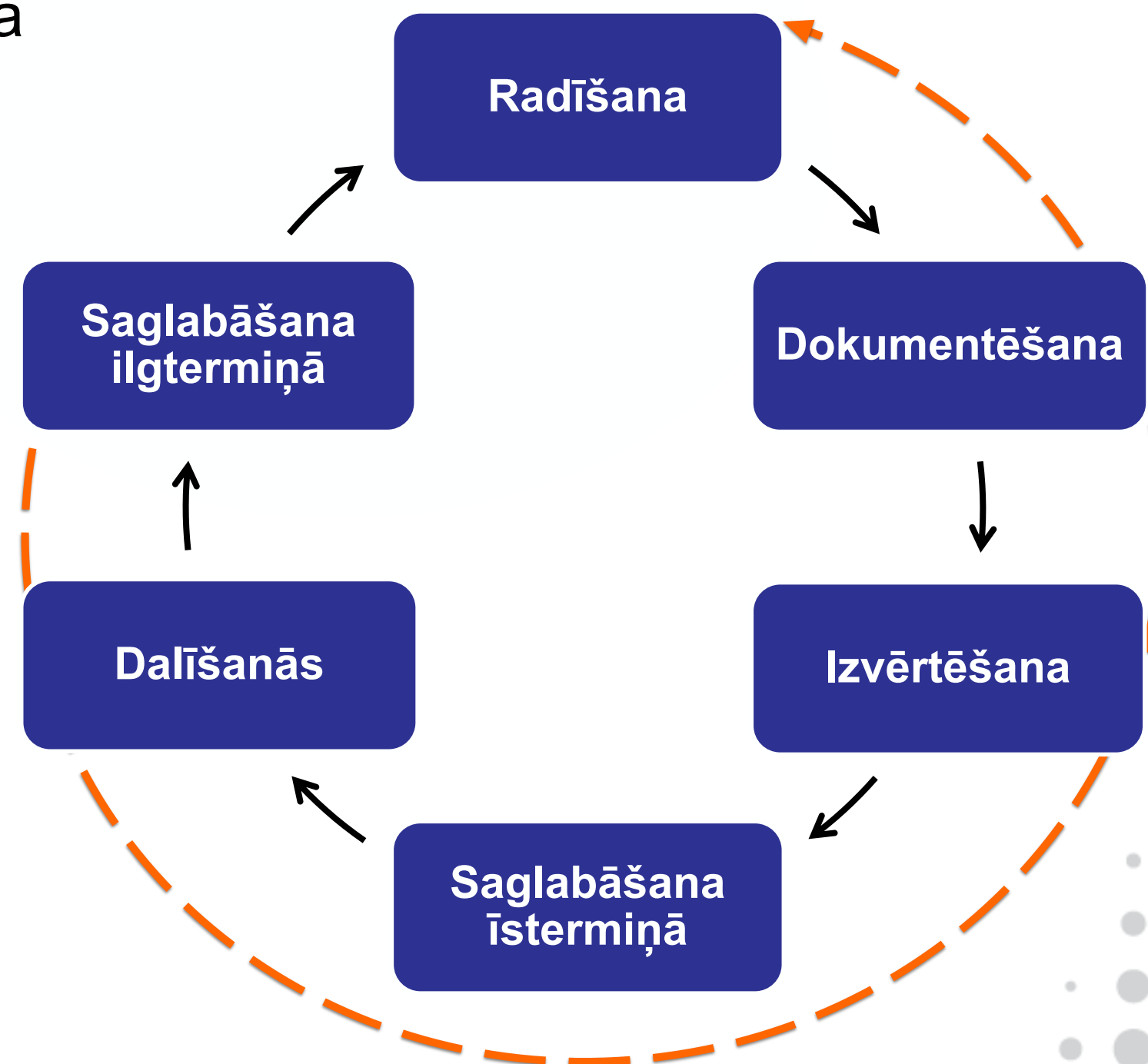
- Datu saglabāšana ilgtermiņā
- Primāri izskatīt iespējas deponēt nozares repozitorijā



Pētniecības datu pārvaldības dzīves cikls (8)

Atvērtās zinātnes principi pētniecībā

- Maina pētniecības ciklu
- Rezultātu publicēšana agrāk un vairāk
- Veicina reproducējamību



Kontaktinformācija

Gita Rozenberga

Nacionālais atvērtās piekļuves dienests

e-pasts: napd@lu.lv

Paldies!

